

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 17:51:35 ; Search time 81 seconds  
(without alignments)  
17.636 Million cell updates/sec

Title: US-09-994-617-4

Perfect score: 41

Sequence: 1 QITEKLEK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	41	100.0	9	22 AAB50528	Anti-HIV peptide P
2	41	100.0	1003	21 AAB69288	HIV-1 non-subtype
3	37	90.2	1893	23 AAU84593	HIV Cassette A2 pr
4	37	90.2	1896	23 AAU84590	HIV Cassette A1 pr
5	37	90.2	5746	23 AAU84596	HIV complete Savin
6	36	87.8	1805	13 AAR27204	Rat nestin. Rattu
7	36	87.8	1805	15 AARG0126	Rat nestin protein
8	35	85.4	49	22 AAG74896	Human colon cancer
9	35	85.4	237	24 AAB726669	Human protein modi

10	35	85.4	256	21 AAB43438	Human cancer assoc
11	35	85.4	415	22 AAB59651	Drosophila melanog
12	34	82.9	16	9 AAB93617	Peptides capable o
13	34	82.9	332	9 AAB80891	Testis-specific la
14	34	82.9	357	21 AAG36241	Arabidopsis thalia
15	34	82.9	424	21 AAG36240	Arabidopsis thalia
16	34	82.9	501	21 AAG62339	Arabidopsis thalia
17	34	82.9	501	21 AAY67413	Arabidopsis aldehy
18	34	82.9	682	22 AAB62036	P. furiosus mcm po
19	34	82.9	1002	9 AAB81861	Sequence encoded b
20	34	82.9	1002	19 AAW72993	HIV isolate LAV.MA
21	34	82.9	1003	11 AAR08060	HIV-1 pol protein
22	34	82.9	1049	22 AAB62034	P. furiosus mcm po
23	33	80.5	9	15 AAY38234	HIV-derived HLA-bi
24	33	80.5	9	20 AAY45806	Immunogenic peptid
25	33	80.5	9	21 AAY66322	HLA-A11-binding HI
26	33	80.5	9	22 AAB14762	HIV A03 super moti
27	33	80.5	9	22 AAB21538	HIV A03 motif pol
28	33	80.5	9	22 AAB23445	HIV A11 motif pol
29	33	80.5	9	22 AAB50527	Anti-HIV peptide p
30	33	80.5	10	24 AAB69776	Human immunodefici
31	33	80.5	10	24 AAB69836	Human immunodefici
32	33	80.5	11	22 AAB14934	HIV A03 super moti
33	33	80.5	11	22 AAB21291	HIV A03 motif pol
34	33	80.5	11	22 AAB21472	HIV A03 motif pol
35	33	80.5	11	22 AAB23254	HIV A11 motif pol
36	33	80.5	11	22 AAB23270	HIV A11 motif pol
37	33	80.5	11	22 AAB23397	HIV A11 motif pol
38	33	80.5	15	19 AAB85184	Helper T-cell pept
39	33	80.5	15	22 AAB24696	HIV DR super motif
40	33	80.5	15	22 AAB24753	HIV DR super motif
41	33	80.5	15	22 AAB24754	HIV DR super motif
42	33	80.5	86	23 AAB33770	Human ORF2743 prot
43	33	80.5	191	21 AAG61742	Arabidopsis thalia
44	33	80.5	247	18 AAW19339	Hybrid polypeptide
45	33	80.5	350	23 AAB90137	Human polypeptide

#### ALIGNMENTS

RESULT 1

AAB50528

ID AAB50528 standard; peptide; 9 AA.

XX AAB50528;

XX 15-MAR-2001 (first entry)

DT Anti-HIV peptide Pol 675-9-5K8E SEQ ID NO:4.

DE Anti-HIV; human immunodeficiency virus; cytotoxic T lymphocyte; CTL;

KW AIDS; infection.

XX Human immunodeficiency virus type 1.

XX WO200073465-A1.

XX 07-DEC-2000.

XX 29-MAY-2000; 2000WO-JP03455.

XX 28-MAY-1999; 95JP-0149589.

XX (AJIN ) AJINOMOTO CO INC.

XX Takiguchi M, Miwa K;

XX WPI; 2001-049939/06.

XX New peptides that induce HIV-specific cytotoxic lymphocytes for treatment and prevention of AIDS

PS Claim 1; Page 26; 31pp; Japanese.

XX AAB50525 to AAB50531 represent peptides which can induce cytotoxic T  
 CC lymphocytes (CTL) specific to HIV-infected cells. The peptides have  
 CC anti-HIV activity. They can be used in the production of vaccines  
 CC against HIV infection, and treatment and prevention of AIDS. The  
 CC peptides are especially effective against HIV infection and AIDS  
 CC in Asian patients.

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 41; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 1 QIIEKLIK 9

# RESULT 2

AA69288  
 ID AAB69288 standard; Protein; 1003 AA.

XX AC AAB69288;

XX 20-APR-2001 (first entry)

XX HIV-1 non-subtype B clone 94CY017-41 pol protein.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env;  
 KW vpr; vif; vpr; tat; rev; nef; vaccine.

XX OS Human immunodeficiency virus type 1.

XX PN WO200026416-A1.

XX 11-MAY-2000.

XX 25-OCT-1999; 99WO-US24837.

XX 02-NOV-1998; 98US-0184418.

XX (UABR-) UAB RES FOUND.

XX Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency  
 PT virus type 1 useful for detecting and treating AIDS comprises a  
 PT specific nucleotide sequence -

XX Claim 41; Fig 15; 131pp; English.

XX The present invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection.

XX SQ Sequence 1003 AA;

Query Match 100.0%; Score 41; DB 21; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 675 QIIEKLIK 683

# RESULT 3

AA84593

ID AA84593 standard; Protein; 1893 AA.

XX AC AA84593;

XX 06-MAY-2002 (first entry)

XX HIV Cassette A2 protein.

XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia.

XX OS Human immunodeficiency virus type 1.  
 OS Synthetic.

XX PN WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI; 2002-147575/19.

XX N-PSDB; ABK36432.

XX New synthetic polypeptides having several different segments of at  
 PT least one parent polypeptide linked together differently compared to  
 PT the linkage in the parent polypeptide, for inducing immune response  
 PT against a pathogen or cancer -

XX Example 1; SEQ ID No 399; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising  
 CC several different segments of at least one parent polypeptide linked  
 CC together in a different relationship relative to their linkage in the  
 CC parent polypeptide to impede, abrogate or otherwise alter at least one  
 CC function associated with the parent polypeptide and for inducing an  
 CC immune response against a pathogen or cancer. Also included are a  
 CC synthetic polynucleotide encoding and a computer system for  
 CC designing the synthetic polypeptides. The synthetic polypeptides and  
 CC polynucleotides are referred to as a Savine. The synthetic polypeptide is  
 CC useful for modulating immune responses preferably directed against a  
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,  
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone  
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.  
 CC Compositions comprising the polypeptide may be used in the treatment or  
 CC prophylaxis against viral (such as infections caused by HIV (human  
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis  
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,  
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is  
 CC a cassette protein consisting of several peptides derived from  
 CC a parent protein. One or more cassettes are used to construct a Savine of  
 CC the invention.

XX SQ Sequence 1893 AA;

Query Match 90.2%; Score 37; DB 23; Length 1893;  
 Best Local Similarity 88.9%; Pred. No. 6e+02;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 OIIEKLIK 9  
 Db 574 QIIEKLIK 582

Best Local Similarity 88.9%; Pred. No. 6e+02;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 OIIEKLIK 9  
 Db 577 QIIEKLIK 585

RESULT 5  
 AAU84596  
 ID AAU84596 standard; Protein; 5746 AA.  
 AC AAU84596;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DE HIV complete Savine protein.  
 XX  
 XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia.  
 XX  
 OS Human immunodeficiency virus type 1.  
 OS Synthetic.  
 XX  
 PN WO200190197-A1.  
 PD 29-NOV-2001.  
 XX  
 XX 25-MAY-2001; 2001WO-AU00622.  
 PF  
 XX 26-MAY-2000; 2000AU-0007761.  
 PR  
 XX (AUSU) UNIV AUSTRALIAN NAT.  
 PA  
 PI Thomson SA, Ramshaw IA;  
 XX  
 XX WPI; 2002-147575/19.  
 DR N-PSDB; ABK36429.  
 XX

New synthetic polypeptides having several different segments of at least one parent polypeptide linked together differently compared to the linkage in the parent polypeptide, for inducing immune response against a pathogen or cancer

Example 1; SEQ ID No 393; 364pp; English.

The invention relates to a new synthetic polypeptide (I) comprising several different segments of at least one parent polypeptide linked together in a different relationship relative to their linkage in the parent polypeptide to impede, abrogate or otherwise alter at least one function associated with the parent polypeptide and for inducing an immune response against a pathogen or cancer. Also included are a synthetic polynucleotide encoding and a computer system for designing the synthetic polypeptides. The synthetic polypeptides and polynucleotides are referred to as a Savine. The synthetic polypeptides are useful for modulating immune responses preferably directed against a pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone liver, oesophagus, brain, testicle, uterus), as potentiating agents. Compositions comprising the polypeptide may be used in the treatment or prophylaxis against viral (such as infections caused by HIV (human immunodeficiency virus), hepatitis, influenza, Japanese encephalitis virus, Epstein-Barr virus and respiratory syncytial virus), bacterial (e.g., infections caused by Neisseria, Meningococcal, Haemophilus, Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic (e.g., infections caused by Plasmodium, Schistosoma, Leishmania, Trypanosoma, Toxoplasma and Giardia) infections. The present invention is a cassette protein consisting of several peptides derived from a parent protein. One or more cassettes are used to construct a savine of the invention.

Query Match 90.2%; Score 37; DB 23; Length 1896;

QY Sequence 1896 AA;  
 XX  
 SQ

RESULT 4  
 AAU84590  
 ID AAU84590 standard; Protein; 1896 AA.  
 AC AAU84590;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DE HIV Cassette A1 protein.  
 XX  
 XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia.  
 XX  
 OS Human immunodeficiency virus type 1.  
 OS Synthetic.  
 XX  
 PN WO200190197-A1.  
 PD 29-NOV-2001.  
 XX  
 XX 25-MAY-2001; 2001WO-AU00622.  
 PF  
 XX 26-MAY-2000; 2000AU-0007761.  
 PR  
 XX (AUSU) UNIV AUSTRALIAN NAT.  
 PA  
 PI Thomson SA, Ramshaw IA;  
 XX  
 XX WPI; 2002-147575/19.  
 DR N-PSDB; ABK36429.  
 XX

New synthetic polypeptides having several different segments of at least one parent polypeptide linked together differently compared to the linkage in the parent polypeptide, for inducing immune response against a pathogen or cancer

Example 1; SEQ ID No 393; 364pp; English.

The invention relates to a new synthetic polypeptide (I) comprising several different segments of at least one parent polypeptide linked together in a different relationship relative to their linkage in the parent polypeptide to impede, abrogate or otherwise alter at least one function associated with the parent polypeptide and for inducing an immune response against a pathogen or cancer. Also included are a synthetic polynucleotide encoding and a computer system for designing the synthetic polypeptides. The synthetic polypeptides and polynucleotides are referred to as a Savine. The synthetic polypeptides are useful for modulating immune responses preferably directed against a pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone liver, oesophagus, brain, testicle, uterus), as potentiating agents. Compositions comprising the polypeptide may be used in the treatment or prophylaxis against viral (such as infections caused by HIV (human immunodeficiency virus), hepatitis, influenza, Japanese encephalitis virus, Epstein-Barr virus and respiratory syncytial virus), bacterial (e.g., infections caused by Neisseria, Meningococcal, Haemophilus, Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic (e.g., infections caused by Plasmodium, Schistosoma, Leishmania, Trypanosoma, Toxoplasma and Giardia) infections. The present invention is a cassette protein consisting of several peptides derived from a parent protein. One or more cassettes are used to construct a savine of the invention.

Query Match 90.2%; Score 37; DB 23; Length 1896;

QY Sequence 1896 AA;  
 XX  
 SQ

RESULT 5  
 AAU84596  
 ID AAU84596 standard; Protein; 5746 AA.  
 AC AAU84596;  
 XX  
 XX 08-MAY-2002 (first entry)  
 DE HIV complete Savine protein.  
 XX  
 XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia.  
 XX  
 OS Human immunodeficiency virus type 1.  
 OS Synthetic.  
 XX  
 PN WO200190197-A1.  
 PD 29-NOV-2001.  
 XX  
 XX 25-MAY-2001; 2001WO-AU00622.  
 PF  
 XX 26-MAY-2000; 2000AU-0007761.  
 PR  
 XX (AUSU) UNIV AUSTRALIAN NAT.  
 PA  
 PI Thomson SA, Ramshaw IA;  
 XX  
 XX WPI; 2002-147575/19.  
 DR N-PSDB; ABK36435.  
 XX

New synthetic polypeptides having several different segments of at least one parent polypeptide linked together differently compared to the linkage in the parent polypeptide, for inducing immune response against a pathogen or cancer

Example 1; Fig 15; 364pp; English.

The invention relates to a new synthetic polypeptide (I) comprising several different segments of at least one parent polypeptide linked together in a different relationship relative to their linkage in the parent polypeptide to impede, abrogate or otherwise alter at least one function associated with the parent polypeptide and for inducing an immune response against a pathogen or cancer. Also included are a synthetic polynucleotide encoding and a computer system for designing the synthetic polypeptides. The synthetic polypeptides and polynucleotides are referred to as a Savine. The synthetic polypeptides are useful for modulating immune responses preferably directed against a pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone liver, oesophagus, brain, testicle, uterus), as potentiating agents. Compositions comprising the polypeptide may be used in the treatment or prophylaxis against viral (such as infections caused by HIV (human immunodeficiency virus), hepatitis, influenza, Japanese encephalitis virus, Epstein-Barr virus and respiratory syncytial virus), bacterial (e.g., infections caused by Neisseria, Meningococcal, Haemophilus, Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic (e.g., infections caused by Plasmodium, Schistosoma, Leishmania, Trypanosoma, Toxoplasma and Giardia) infections. The present invention is a cassette protein consisting of several peptides derived from a parent protein. One or more cassettes are used to construct a savine of the invention.

Query Match 90.2%; Score 37; DB 23; Length 1896;

QY Sequence 1896 AA;  
 XX  
 SQ

Query Match 90.2%; Score 37; DB 23; Length 5746;  
 Best Local Similarity 88.9%; Pred. No. 1.8e+03;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
 Db 574 QIIEKLIK 582

## RESULT 6

AAR27204  
 ID AAR27204 standard; Protein; 1805 AA.

XX AC AAR27204;

XX 25-MAR-2003 (updated)

DT 20-MAY-1998 (first entry)

XX DE Rat nestin.

XX Intermediate filament; central nervous system; brain tumour;  
 neurofilament.

XX OS Rattus norvegicus.

XX WO9214821-A2.

XX 03-SEP-1992.

XX 21-FEB-1992; 92WO-US01375.

XX 22-FEB-1991; 91US-0660412.

XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX Lendahl U, McKay RDG;

XX WPI; 1992-316175/38.

XX N-PSDB; AAQ28398.

XX Diagnosis of pre-disposition to brain tumours - using DNA  
 encoding nestin which distinguishes neural multi-potential stem  
 cells from neuronal, glial and muscle cells

XX Claim 3; Fig 1; 63pp; English.

XX The amino acid sequence of the nestin gene which was deduced from  
 the nucleotide sequence, suggests that nestin is a member of the  
 intermediate filament protein family. The rat nestin amino acid  
 sequence shows 75% similarity with the human nestin sequence. There  
 is more than 60% identity between the two sequences. Antibodies to  
 nestin protein can be used in vivo diagnosis of brain tumours.  
 See also AAQ28399 for the human nestin gene.  
 (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 1805 AA;

Query Match 87.8%; Score 36; DB 13; Length 1805;  
 Best Local Similarity 77.8%; Pred. No. 8.5e+02;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9

Db 713 QILERLIK 721

## RESULT 7

AAR60126

ID AAR60126 standard; Protein; 1805 AA.

XX AC AAR60126;

XX 25-MAR-2003 (updated)

DT 21-MAR-1995 (first entry)

XX DE Rat nestin protein is useful to identify brain tumours.

XX KW nestin gene; brain tumour; neoplastic cells; glial; neuronal;  
 muscle; neural multipotential stem cell; mammalian brain; detection;  
 diagnosis; medulloblastoma; glioblastoma; oligodendroglioma.

OS Rattus rattus.

XX Key Location/Qualifiers

FT Misc-difference 167..168

FT /note= "encoded by CACGG"

FT Misc-difference 172..173

FT /note= "encoded by CCG G GCC"

XX US5338839-A.

XX 16-AUG-1994.

XX 19-MAR-1992; 92US-0853913.

XX 12-APR-1988; 88US-0180548.

XX 02-JUN-1988; 88US-0201762.

XX 25-OCT-1990; 90US-0603803.

XX 22-FEB-1991; 91US-0660412.

XX 19-MAR-1992; 92US-0853913.

XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX Lendahl U, McKay RDG;

XX WPI; 1994-263332/32.

XX N-PSDB; AAQ70447.

XX Nucleotide and protein sequences for human and rat nestin -  
 distinguishes neural multipotential stem cells and brain tumour  
 cells from more differentiated cell types; for use in the  
 diagnosis of brain tumours

XX Claim 4; Column 35-44; 45pp; English.

XX This sequence is the rat nestin protein encoded by AAQ70447. Its  
 degree of sequence homology to intermediate filaments in the core  
 domain, its dissimilarity in the head and tail domains, and its  
 different splicing pattern suggest that it defines a new class of  
 intermediate filament protein. Nestin protein expression distinguishes  
 neural multipotential stem cells and brain tumour cells from the more  
 differentiated neural cell types (eg., neuronal, glial and muscle  
 cells of the adult brain). The nestin protein can be used in diagnosing  
 tumours of the brain, such as medulloblastomas, glioblastomas and  
 oligodendroglioma. (See also AAR60127).  
 (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 1805 AA;

Query Match 87.8%; Score 36; DB 15; Length 1805;  
 Best Local Similarity 77.8%; Pred. No. 8.5e+02;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9

Db 713 QILERLIK 721

## RESULT 8

AAG74896

ID AAG74896 standard; Protein; 49 AA.

XX AC AAG74896;

XX 03-SEP-2001 (first entry)



DE Human colon cancer antigen protein SEQ ID NO:5660.  
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
PI N-PSDB; AAH3430L.  
DR WPI; 2001-235357/24.  
DR N-PSDB; AAH3430L.  
XX  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
PS Claim 11; Page 7204-7206; 9803pp; English.  
XX  
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAH77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 49 AA;  
Query Match 85.4%; Score 35; DB 22; Length 49;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QIIIEKLIEK 9  
Db ::::|  
13 EVLEKLIEK 21  
RESULT 9  
ABJ26669  
ID ABJ26669 standard; Protein; 237 AA.  
XX  
AC ABJ26669;  
XX  
DT 01-MAY-2003 (first entry)  
XX  
DE Human protein modification + maintenance molecule protein SEQ ID NO 23.  
XX  
KW Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
KW cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
KW antiulcer; hepatotropic; gynaecological; antibacterial; virucide;  
KW protozoacide; antiparasitic; cell proliferative disease; PMOD;

KW protein modification and maintenance molecule; immunogenic fragment;  
KW cancer; autoimmune; inflammatory disease; neurological disorder;  
KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KW protein-protein interaction; drug-target interaction;  
KW gene expression profile; human.  
XX  
OS Homo sapiens.  
XX  
PN WO2003000844-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 18-JUN-2002; 2002WO-US19360.  
XX  
PR 22-JUN-2001; 2001US-300508P.  
PR 06-JUL-2001; 2001US-303445P.  
PR 13-JUL-2001; 2001US-305405P.  
PR 09-AUG-2001; 2001US-311442P.  
PR 24-AUG-2001; 2001US-314821P.  
PR 29-AUG-2001; 2001US-315992P.  
PR 03-MAY-2002; 2002US-378205P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;  
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA;  
PI Yue H, Forsythe IJ, Barroso I, Ramkumar J, Griffin JA, Li JX;  
PI Yang J, Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML;  
PI Yao MG, Walla NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY;  
PI Tran UK, Elliott VS, Luo W, Sprague MW, Tang YF, Lu Y;  
PI Zebajradjan Y;  
XX  
DR WPI; 2003-184039/18.  
DR N-PSDB; ABT23222.  
XX  
PT New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis  
PT and infections -  
XX  
PS Claim 78; Page 197; 225pp; English.  
XX  
CC The invention relates to an isolated polypeptide comprising: any of 28  
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This sequence represents a human PMOD protein  
CC of the invention.  
XX  
SQ Sequence 237 AA;  
Query Match 85.4%; Score 35; DB 24; Length 237;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
 ID AAB43438 standard; Protein; 256 AA.  
 Db 170 EVLEKLIK 178

RESULT 10  
 AAB43438  
 AC AAB43438;  
 XX  
 XX  
 DT 08-FEB-2001 (first entry)  
 DE Human cancer associated protein sequence SEQ ID NO:883.  
 DE Human: cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;  
 KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;  
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200055350-AL.  
 PN  
 XX  
 PD 21-SEP-2000.  
 XX  
 XX 08-MAR-2000; 2000WO-US05882.  
 XX  
 XX 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 XX WPI; 2000-587533/55.  
 DR N-PSDB; AAC77647.  
 XX  
 XX Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer.  
 PT  
 XX  
 XX Claim 11; Page 1438-1439; 2352pp; English.  
 XX  
 XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnery; immunomodulator;  
 CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 256 AA;

Query Match 85.4%; Score 35; DB 21; Length 256;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
 Db 189 EVLEKLIK 197

RESULT 11  
 ABB59651  
 ID ABB59651 standard; Protein; 415 AA.  
 XX  
 AC ABB59651;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 5745.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW Drosophila; development;  
 KW pharmacological.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PR  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL03754.  
 DR  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 XX Disclosure; SEQ ID NO 5745; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 415 AA;

Query Match 85.4%; Score 35; DB 22; Length 415;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
 Db 262 KIVEKLUK 270

RESULT 12  
 AAP82617  
 ID AAP82617 standard; peptide; 16 AA.  
 XX  
 AC AAP82617;

```

XX 25-MAR-2003 (updated)
DT 30-NOV-1990 (first entry)
XX
DE Peptides capable of inducing prodn. of antibodies to human semen
DE lactate dehydrogenase-C4.
XX
KW Semen lactate dehydrogenase-C4; antibodies; female fertility; vaccines.
XX
OS synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 2..16
FT Peptide 3..16
FT Peptide 4..16
FT Peptide 5..16
XX
PN EP270055-A.
XX
PD 08-JUN-1988.
XX
PF 30-NOV-1987; 87EP-0117687.
XX
PR 01-DEC-1986; 86US-0936170.
XX
PA (NOUN ) UNIV NORTHWESTERN.
XX
PI Goldberg E, Millan JL;
XX
WPI; 1988-156289/23.
XX
New peptide(s) producing antibodies to human semen enzyme -
useful for prodn of antifertility vaccines.
XX
PS Claim 1; Page 8; 9pp; English.
XX
CC This peptide and fragments are useful in a vaccine compsn. for
CC reducing female fertility. It corresp. to an antigenic region
CC of semen lactate dehydrogenase (LDH)-C4 and therefore induces
CC prodn. of antibodies against this. See also AAP82667-71 and AAN82189.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 16 AA;
Query Match 82.9%; Score 34; DB 9; Length 16;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 QIIEKLIE 8
DB 6 QLIIEKLIE 13
XX
RESULT 13
AAP80891
ID AAP80891 standard; protein; 332 AA.
XX
AC AAP80891;
XX
DT 25-MAR-2003 (updated)
DT 29-NOV-1990 (first entry)
XX
DE Testis-specific lactate dehydrogenase subunit LDH-C4.
XX
KW Testis-specific lactate dehydrogenase; vaccine; female fertility.
XX
OS synthetic.
XX
PN EP270056-A.
XX
PD 08-JUN-1988.
XX
PF 30-NOV-1987; 87EP-0117688.

```

```

XX 01-DEC-1986; 86US-0936451.
PR 12-JUN-1987; 87US-0061894.
XX
PA (NOUN ) UNIV NORTHWESTERN.
XX
PI Goldberg E, Millan JL;
XX
WPI; 1988-156290/23.
DR N-PSDB; AAN80494.
XX
New recombinant DNA encoding sub-unit of testis lactate hydrogenase -
expressing antigenic polypeptide useful in vaccines for reducing
female fertility.
XX
PS Disclosure; Page ?; ?pp; English.
XX
CC This testis-specific lactate dehydrogenase sub-unit (LDH-C4) or one
CC or more antigenic regions can be used in a vaccine prepn. to reduce
CC female fertility. Larger peptides (or fusion proteins) are antigenic
CC and smaller peptides become so when coupled to e.g. tetanus toxoid.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 332 AA;
Query Match 82.9%; Score 34; DB 9; Length 332;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 QIIEKLIE 8
DB 7 QLIIEKLIE 14
XX
RESULT 14
AAG36241
ID AAG36241 standard; Protein; 357 AA.
XX
AC AAG36241;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44385.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

```

PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0132407.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.9%; Score 34; DB 21; Length 357;
Best Local Similarity 55.6%; Pred. No. 3.8e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 QIIKLEIK 9
Db      173 KVKELVEK 181

RESULT 15
AAG36240
ID AAG36240 standard; Protein; 424 AA.
XX      AC
XX      AAG36240;
DT 18-OCT-2000 (first entry)
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 44384.
XX      Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX      Arabidopsis thaliana.
XX      EP1033405-A2.
PD 06-SEP-2000.
XX      PF
XX      25-FEB-2000; 2000EP-0301439.
XX      25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.

99US-0134218.
99US-0134219.
99US-0134221.
99US-0134370.
99US-0134768.
99US-0134941.
99US-0135124.
99US-0135353.
99US-0135629.
99US-0136021.
99US-0136392.
99US-0136782.
99US-0137222.
99US-0137528.
99US-0137502.
99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139492.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139763.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
99US-0143624.
99US-0144005.
99US-0144085.
99US-0144086.
99US-0144325.
99US-0144331.
99US-0144332.
99US-0144333.
99US-0144334.
99US-0144335.
99US-0144352.
99US-0144632.
99US-0144884.
99US-0144814.
99US-0145086.
99US-0145088.
99US-0145085.
99US-0145087.
99US-0145089.
99US-0145192.
99US-0145145.
99US-0145218.

```

PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149802.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 82.9%; Score 34; DB 21; Length 424;  
Best Local Similarity 55.6%; Pred. No. 4.5e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9  
Db 240 KVEKLEK 248

Search completed: July 29, 2003, 18:21:06  
Job time : 83 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 18:19:46 ; Search time 29 Seconds  
(without alignments)  
13.131 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIEKLIK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	90.2	493	US-09-328-352-6753	Sequence 6753, Ap
2	36	87.8	1805	US-07-853-913-2	Sequence 2, Appli
3	34	82.9	327	US-08-748-068-2	Sequence 2, Appli
4	34	82.9	1003	US-07-743-357-9	Sequence 9, Appli
5	33	80.5	9	US-08-159-339A-527	Sequence 527, App
6	33	80.5	15	US-09-009-953-86	Sequence 86, Appli
7	33	80.5	560	US-09-752-652-1	Sequence 1, Appli
8	33	80.5	562	US-09-117-217-14	Sequence 14, Appli
9	33	80.5	562	US-09-735-487-14	Sequence 14, Appli
10	33	80.5	729	US-09-107-532A-3970	Sequence 3970, Ap
11	33	80.5	913	US-07-743-357-6	Sequence 6, Appli
12	33	80.5	913	US-07-743-357-22	Sequence 22, Appli
13	33	80.5	1003	US-07-743-357-8	Sequence 8, Appli
14	33	80.5	1003	US-07-743-357-10	Sequence 10, Appli
15	33	80.5	1003	US-09-309-572-17	Sequence 17, Appli
16	33	80.5	1004	US-07-743-357-7	Sequence 7, Appli
17	33	80.5	1005	US-07-743-357-1	Sequence 1, Appli
18	33	80.5	1014	US-09-319-588C-6	Sequence 6, Appli
19	33	80.5	1015	US-08-463-210-9	Sequence 9, Appli
20	33	80.5	1015	US-09-124-900-3	Sequence 3, Appli
21	33	80.5	1016	US-07-743-357-2	Sequence 2, Appli
22	33	80.5	1016	US-07-743-357-3	Sequence 3, Appli
23	33	80.5	1016	US-07-743-357-4	Sequence 4, Appli
24	33	80.5	1016	US-07-743-357-5	Sequence 5, Appli
25	32	78.0	27	US-08-449-500-82	Sequence 82, Appli
26	32	78.0	27	US-08-449-317A-82	Sequence 82, Appli
27	32	78.0	27	US-08-477-022-82	Sequence 82, Appli

28	32	78.0	27	2	US-08-449-447-82	Sequence 82, Appli
29	32	78.0	27	2	US-08-184-328-82	Sequence 82, Appli
30	32	78.0	27	2	US-08-521-097-82	Sequence 82, Appli
31	32	78.0	28	1	US-08-449-500-81	Sequence 81, Appli
32	32	78.0	28	1	US-08-449-317A-81	Sequence 81, Appli
33	32	78.0	28	2	US-08-477-022-81	Sequence 81, Appli
34	32	78.0	28	2	US-08-449-447-81	Sequence 81, Appli
35	32	78.0	28	2	US-08-184-328-81	Sequence 81, Appli
36	32	78.0	28	2	US-08-521-097-81	Sequence 81, Appli
37	32	78.0	32	1	US-08-449-500-56	Sequence 56, Appli
38	32	78.0	32	1	US-08-449-500-80	Sequence 80, Appli
39	32	78.0	32	1	US-08-449-317A-56	Sequence 56, Appli
40	32	78.0	32	1	US-08-449-317A-80	Sequence 80, Appli
41	32	78.0	32	2	US-08-477-022-56	Sequence 56, Appli
42	32	78.0	32	2	US-08-477-022-80	Sequence 80, Appli
43	32	78.0	32	2	US-08-449-447-56	Sequence 56, Appli
44	32	78.0	32	2	US-08-449-447-80	Sequence 80, Appli
45	32	78.0	32	2	US-08-184-328-56	Sequence 56, Appli

ALIGNMENTS

RESULT 1

US-09-328-352-6753  
; Sequence 6753, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6753  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6753

Query Match 90.2%; Score 37; DB 4; Length 493;  
Best Local Similarity 88.9%; Pred. NO. 31;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QIEKLIK 9  
Db 310 QFIEKLIK 318

RESULT 2

US-07-853-913-2  
; Sequence 2, Application US/07853913  
; Patent No. 5338839  
; GENERAL INFORMATION:  
; APPLICANT: McKay, Ronald D.G.  
; APPLICANT: Lendahl, Urban  
; TITLE OF INVENTION: Nestin Expression as an Indicator of  
; TITLE OF INVENTION: Neuroepithelial Tumors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-853-913-2

Query Match      87.88; Score 36; DB 1; Length 1805;
Best Local Similarity 77.88; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 OIIEKLIK 9
      11:11111
Db      713 OILERLIK 721

RESULT 3
US-08-748-068-2
; Sequence 2, Application US/08748068
; Patent No. 5770410
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Chiral Synthesis
; NUMBER OF SEQUENCES: 15
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,068
; FILING DATE: 12-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/256,959
; FILING DATE: 03-OCT-1994
; APPLICATION NUMBER: GB 92 02033.8
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92 04702.6
; FILING DATE: 04-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 93/00204
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO

; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (16~17)
; OTHER INFORMATION: /note= "--- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (27~28)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (40~41)
; OTHER INFORMATION: /note= "--- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (41~42)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (42~43)
; OTHER INFORMATION: /note= "s--- numbering
; OTHER INFORMATION: discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (52~53)
; OTHER INFORMATION: /note= "--- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (81~82)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (85~86)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (95~96)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (127~128)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (150~151)
; OTHER INFORMATION: /note= "---- numbering
; OTHER INFORMATION: discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (188~189)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (193~194)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (208~209)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (243~244)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (247~248)
; OTHER INFORMATION: /note= "- numbering discontinuity"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: (259~260)
; OTHER INFORMATION: /note= "--- numbering discontinuity"
; FEATURE:
```



NAME/KEY: Modified-site  
LOCATION: (262~263)  
OTHER INFORMATION: /note= "--- numbering discontinuity"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: (275~276)  
OTHER INFORMATION: /note= "---- numbering  
OTHER INFORMATION: discontinuity"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: (298~299)  
OTHER INFORMATION: /note= "- numbering discontinuity"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: (308~309)  
OTHER INFORMATION: /note= "- numbering discontinuity"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: (326~327)  
OTHER INFORMATION: /note= "- numbering discontinuity"  
US-08-748-068-2

Query Match 82.9%; Score 34; DB 1; Length 327;  
Best Local Similarity 87.5%; Pred. No. 71;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIKLEIE 8  
Db 6 QIIKLEIE 13

## RESULT 4

US-07-743-357-9  
; Sequence 9, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743.357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1003 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human Immunodeficiency virus type 1  
; STRAIN: MAL  
; US-07-743-357-9

Query Match 82.9%; Score 34; DB 2; Length 1003;  
Best Local Similarity 77.8%; Pred. No. 21e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIKLEIE 9  
Db 674 QIIKLEIE 682

## RESULT 5

US-08-159-339A-527  
; Sequence 527, Application US/08159339A  
; Patent No. 60371135  
; GENERAL INFORMATION:  
; APPLICANT: Kubo, Ralph. T.  
; APPLICANT: Grey, Howard M.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Celis, Esteban  
; TITLE OF INVENTION: HLA Binding peptides and Their  
; TITLE OF INVENTION: Uses  
; NUMBER OF SEQUENCES: 1254  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,339A  
; FILING DATE: 29-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,666  
; FILING DATE: 07-AUG-1992  
; APPLICATION NUMBER: US 08/027,746  
; FILING DATE: 05-MAR-1993  
; APPLICATION NUMBER: US 08/103,396  
; FILING DATE: 06-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 018623-005030US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 527:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-159-339A-527

Query Match 80.5%; Score 33; DB 3; Length 9;  
Best Local Similarity 77.8%; Pred. No. 2.5e+05;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIKLEIE 9

```
Db      1 QIIEQLIKK 9
||||:|:|
;
; CURRENT APPLICATION NUMBER: US/09/752.652
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/167,940
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 100
; OTHER INFORMATION: Xaa-Leu or Ile
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 103
; OTHER INFORMATION: Xaa-Lys or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 106
; OTHER INFORMATION: Xaa=Val, Ile or Ala
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 108
; OTHER INFORMATION: Xaa=Val or Ile
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 138
; OTHER INFORMATION: Xaa=Glu or Lys
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 181
; OTHER INFORMATION: Xaa-Tyr or Cys
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: Xaa at position 188 can be Tyr or Cys
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: Xaa at position 236 can be Pro or Leu
; US-09-752-652-1
;
; Query Match      80.5%; Score 33; DB 4; Length 560;
; Best Local Similarity 77.8%; Pred. No. 1.8e+02;
; Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QIIEKLIK 9
||||:|:|
Db      520 QIIEQLIKK 528

RESULT 8
US-09-117-217-14
; Sequence 14, Application US/09117217
; Patent No. 6221578
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOGS, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; TITLE OF INVENTION: OF HUMAN HIV STRAINS
; FILE REFERENCE: 1377-125p
; CURRENT APPLICATION NUMBER: US/09/117,217
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 562
; TYPE: PRT
; ORGANISM: HIV-HXB2

Db      1 QIIEQLIKK 9
||||:|:|
;
; CURRENT APPLICATION NUMBER: US/09/752.652
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/167,940
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 100
; OTHER INFORMATION: Xaa-Leu or Ile
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 103
; OTHER INFORMATION: Xaa-Lys or Asn
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 106
; OTHER INFORMATION: Xaa=Val, Ile or Ala
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 108
; OTHER INFORMATION: Xaa=Val or Ile
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 138
; OTHER INFORMATION: Xaa=Glu or Lys
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 181
; OTHER INFORMATION: Xaa-Tyr or Cys
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: Xaa at position 188 can be Tyr or Cys
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (0)...(0)
; OTHER INFORMATION: Xaa at position 236 can be Pro or Leu
; US-09-752-652-1
;
; Query Match      80.5%; Score 33; DB 4; Length 560;
; Best Local Similarity 77.8%; Pred. No. 1.8e+02;
; Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QIIEKLIK 9
||||:|:|
Db      520 QIIEQLIKK 528

RESULT 7
US-09-752-652-1
; Sequence 1, Application US/09752652
; Patent No. 6503123
; GENERAL INFORMATION:
; APPLICANT: Roberts, Grace B.
; APPLICANT: Furfine, Eric S.
; APPLICANT: Porter, David, J.T.
; TITLE OF INVENTION: CONTINUOUS TIME RESOLVED RESONANCE
; TITLE OF INVENTION: ENERGY TRANSFER ASSAY FOR POLYNUCLEIC ACID POLYMERASE
; TITLE OF INVENTION: ACTIVITY
; FILE REFERENCE: P03761
```

## US-09-117-217-14

Query Match 80.5%; Score 33; DB 3; Length 562;  
Best Local Similarity 77.8%; Pred No. 1.8e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
||||:|:|  
Db 520 QIIEQLIK 528

## RESULT 9

US-09-735-487-14  
; Sequence 14, Application US/09735487  
; Patent No. 6528251  
; GENERAL INFORMATION:  
; APPLICANT: de BETHUNE, Marie-Pierre  
; APPLICANT: HERTOGS, Kurt  
; APPLICANT: PAUWELS, Rudi  
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE  
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY  
; TITLE OF INVENTION: OF HUMAN HIV STRAINS  
; FILE REFERENCE: 1377-125P  
; CURRENT APPLICATION NUMBER: US/09/735,487  
; CURRENT FILING DATE: 2000-12-14  
; PRIOR APPLICATION NUMBER: 09/117,217  
; PRIOR FILING DATE: 1998-07-24  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: HIV-HXB2  
US-09-735-487-14

Query Match 80.5%; Score 33; DB 4; Length 562;  
Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
||||:|:|  
Db 520 QIIEQLIK 528

## RESULT 10

US-09-107-532A-3970  
; Sequence 3970, Application US/09107532A  
; Patent No. 6363275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 3970:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 729 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...729  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3970:  
US-09-107-532A-3970

Query Match 80.5%; Score 33; DB 4; Length 729;  
Best Local Similarity 88.9%; Pred. No. 2.3e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
|:|:|:|:|  
Db 165 QGIEKLIK 173

## RESULT 11

US-07-743-357-6  
; Sequence 6, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 913 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein

;  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: MN  
; US-07-743-357-6

Query Match 80.5%; Score 33; DB 2; Length 913;  
Best Local Similarity 77.8%; Pred. No. 2.9e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 584 QIIEQLIKK 592

## RESULT 12

US-07-743-357-22  
; Sequence 22, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: KIM 1H8

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045

INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 913 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: HXB2  
; US-07-743-357-22

Query Match 80.5%; Score 33; DB 2; Length 913;  
Best Local Similarity 77.8%; Pred. No. 2.9e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 584 QIIEQLIKK 592

## RESULT 13

US-07-743-357-8  
; Sequence 8, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: KIM 1H8

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045

INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1003 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: RF  
; US-07-743-357-8

Query Match 80.5%; Score 33; DB 2; Length 1003;  
Best Local Similarity 77.8%; Pred. No. 3.2e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 674 QIIEQLIKK 682

## RESULT 14

US-07-743-357-10  
; Sequence 10, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological  
; activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario

Search completed: July 29, 2003, 18:24:27  
Job time : 30 secs

COUNTRY: Canada  
ZIP: KIM IHS  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/743,357  
FILING DATE: 21-AUG-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA90/00062  
FILING DATE: 23-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gale, Edwin J.  
REGISTRATION NUMBER: 28,584  
REFERENCE/DOCKET NUMBER: 30924-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613) 237-6900  
TELEFAX: (613) 237-0045  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1003 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Human immunodeficiency virus type 1  
STRAIN: ELI  
US-07-743-357-10

Query Match 80.5%; Score 33; DB 2; Length 1003;  
Best Local Similarity 77.8%; Pred. No. 3.2e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
Db 674 QIIEQLIKK 682

RESULT 15  
US-09-309-572-17  
Sequence 17, Application US/09309572  
Patent No. 6440730  
GENERAL INFORMATION:  
APPLICANT: Heinrich-Pette-Institut  
TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
FILE REFERENCE: P50489  
CURRENT APPLICATION NUMBER: US/09/309,572  
CURRENT FILING DATE: 1999-05-11  
EARLIER APPLICATION NUMBER: DE 198 56 463  
EARLIER FILING DATE: 1998-11-26  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 17  
LENGTH: 1003  
TYPE: PRT  
ORGANISM: Human immunodeficiency virus type 1  
FEATURE:  
OTHER INFORMATION: pol polyprotein  
US-09-309-572-17

Query Match 80.5%; Score 33; DB 4; Length 1003;  
Best Local Similarity 77.8%; Pred. No. 3.2e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
Db 675 QIIEQLIKK 683



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 18:23:16 ; Search time 51 Seconds  
(without alignments)  
20.958 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIIEKLIK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues  
Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US05\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	9	11 US-09-994-617-4	Sequence 4, Appli
2	35	85.4	456	15 US-10-106-698-5670	Sequence 5670, Ap
3	35	85.4	256	9 US-09-925-301-883	Sequence 883, App
4	34	82.9	501	10 US-09-344-882-22	Sequence 22, Appl
5	34	82.9	501	15 US-10-293-865-22	Sequence 22, Appl
6	33	80.5	9	11 US-09-994-617-3	Sequence 3, Appli
7	33	80.5	9	15 US-10-133-210-91	Sequence 29, Appl
8	33	80.5	546	8 US-08-808-031A-29	Sequence 91, Appl
9	33	80.5	560	15 US-10-205-641-1	Sequence 1, Appli
10	33	80.5	561	15 US-10-059-271-87	Sequence 87, Appl
11	33	80.5	562	9 US-09-735-487-14	Sequence 14, Appl
12	33	80.5	850	11 US-09-952-060-2	Sequence 2, Appli
13	33	80.5	850	11 US-09-952-060-4	Sequence 4, Appli
14	33	80.5	851	15 US-10-059-271-88	Sequence 88, Appl
15	33	80.5	875	11 US-09-952-060-6	Sequence 6, Appli

16	33	80.5	875	11	US-09-952-060-8	Sequence 8, Appli
17	33	80.5	1350	11	US-09-952-060-35	Sequence 35, Appl
18	33	80.5	2697	10	US-09-961-527A-5	Sequence 5, Appl1
19	32	78.0	24	11	US-09-843-221A-129	Sequence 129, App
20	32	78.0	24	11	US-09-843-221A-139	Sequence 139, Appl
21	32	78.0	28	11	US-09-843-221A-93	Sequence 93, Appl
22	32	78.0	28	11	US-09-843-221A-103	Sequence 103, App
23	32	78.0	30	11	US-09-843-221A-124	Sequence 124, App
24	32	78.0	30	11	US-09-843-221A-134	Sequence 134, App
25	32	78.0	30	11	US-09-843-221A-145	Sequence 145, App
26	32	78.0	30	11	US-09-843-221A-147	Sequence 147, App
27	32	78.0	30	11	US-09-843-221A-148	Sequence 148, App
28	32	78.0	30	11	US-09-843-221A-149	Sequence 149, App
29	32	78.0	30	11	US-09-843-221A-150	Sequence 150, App
30	32	78.0	30	11	US-09-843-221A-158	Sequence 158, App
31	32	78.0	30	11	US-09-843-221A-159	Sequence 159, App
32	32	78.0	33	11	US-09-843-221A-146	Sequence 146, App
33	32	78.0	34	9	US-09-169-786-6	Sequence 6, Appli
34	32	78.0	34	9	US-09-169-786-7	Sequence 7, Appl1
35	32	78.0	34	9	US-09-169-786-8	Sequence 8, Appl1
36	32	78.0	34	11	US-09-843-221A-88	Sequence 88, Appl
37	32	78.0	34	11	US-09-843-221A-98	Sequence 98, Appl
38	32	78.0	34	11	US-09-843-221A-109	Sequence 109, App
39	32	78.0	34	11	US-09-843-221A-111	Sequence 111, App
40	32	78.0	34	11	US-09-843-221A-112	Sequence 112, App
41	32	78.0	34	11	US-09-843-221A-113	Sequence 113, App
42	32	78.0	34	11	US-09-843-221A-114	Sequence 114, App
43	32	78.0	34	11	US-09-843-221A-122	Sequence 122, App
44	32	78.0	34	11	US-09-843-221A-123	Sequence 123, App
45	32	78.0	37	11	US-09-843-221A-110	Sequence 110, App

## ALIGNMENTS

## RESULT 1

US-09-994-617-4  
; Sequence 4, Application US/09994617  
; Publication No. US20030114369A1  
; GENERAL INFORMATION:  
; APPLICANT: MIWA, Kiyoshi  
; TITLE OF INVENTION: HIV-SPECIFIC CTL INDUCING PEPTIDES AND MEDICAMENTS FOR PREVENT  
; TITLE OF INVENTION: TREATING AIDS COMPRISING THE PEPTIDES  
; FILE REFERENCE: 216457USOCNT  
; CURRENT APPLICATION NUMBER: US/09/994,617  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/03455  
; PRIOR FILING DATE: 2000-05-29  
; PRIOR APPLICATION NUMBER: JP11-149589  
; PRIOR FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 4  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-994-617-4

Query Match 100.0%; Score 41; DB 11; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
DB 1 QIIEKLIK 9

## RESULT 2

US-10-106-698-5670  
; Sequence 5670, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:

```
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5670
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5670

Query Match      85.4%; Score 35; DB 15; Length 49;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9
Db 13 EVLEKLEIK 21

RESULT 3
US-09-925-301-883
; Sequence 883, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05982
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 883
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-883

Query Match      85.4%; Score 35; DB 9; Length 256;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9
Db 189 EVLEKLEIK 197

RESULT 4
US-09-344-882-22
; Sequence 22, Application US/09344882
; Patent No. US20020162137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
US-09-344-882-22

Query Match      82.9%; Score 34; DB 15; Length 501;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9
Db 317 KVVEKLEIK 325

RESULT 5
US-10-293-865-22
; Sequence 22, Application US/10293865
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 217113
; CURRENT APPLICATION NUMBER: US/10/293,865
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/344,882
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-22

Query Match      82.9%; Score 34; DB 15; Length 501;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9
Db 317 KVVEKLEIK 325

RESULT 6
US-09-994-617-3
; Sequence 3, Application US/09994617
; Publication No. US20030114369A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-22
```

```
Query Match      82.9%; Score 34; DB 10; Length 501;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 QIIEKLEIK 9
Db 317 KVVEKLEIK 325
```

```
RESULT 5
US-10-293-865-22
; Sequence 22, Application US/10293865
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 217113
; CURRENT APPLICATION NUMBER: US/10/293,865
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/344,882
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-22
```

```
Query Match      82.9%; Score 34; DB 15; Length 501;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 QIIEKLEIK 9
Db 317 KVVEKLEIK 325
```

```
RESULT 6
US-09-994-617-3
; Sequence 3, Application US/09994617
; Publication No. US20030114369A1
; GENERAL INFORMATION:
```



```
; APPLICANT: TAKIGUCHI, Masafumi
; APPLICANT: MIWA, Kiyoshi
; TITLE OF INVENTION: HIV-SPECIFIC CTL INDUCING PEPTIDES AND MEDICAMENTS FOR PREVENTING
; FILE OF INVENTION: TREATING AIDS COMPRISING THE PEPTIDES
; FILE REFERENCE: 216457US00CNT
; CURRENT APPLICATION NUMBER: US/09/994,617
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: PCT/JPO0/03455
; PRIOR FILING DATE: 2000-05-29
; PRIOR APPLICATION NUMBER: JP11-149589
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-994-617-3

Query Match      80.5%; Score 33; DB 11; Length 9;
Best Local Similarity 77.8%; Pred. No. 4e+05;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 QIIEKLIK 9
      111111111
Db      1 QIIEELIK 9

RESULT 7
US-10-133-210-91
; Sequence 91, Application US/10133210
; Publication No. US20030103964A1
; GENERAL INFORMATION:
; APPLICANT: DeLisi, Charles
; APPLICANT: Berzofsky, Jay
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Vaccaro, Dennis
; APPLICANT: Wang, Zhiping
; APPLICANT: Zhang, Chao
; TITLE OF INVENTION: METHODS FOR DESIGNING MOLECULAR CONJUGATES AND
; FILE REFERENCE: BU-035AX
; CURRENT APPLICATION NUMBER: US/10/133,210
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 281
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-133-210-91

Query Match      80.5%; Score 33; DB 15; Length 9;
Best Local Similarity 77.8%; Pred. No. 4e+05;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 QIIEKLIK 9
      111111111
Db      1 QIIEELIK 9

RESULT 8
US-08-808-031A-29
; Sequence 29, Application US/08808031A
; Publication No. US20020048802A1
; GENERAL INFORMATION:
; APPLICANT: Inouye, Sumiko
; APPLICANT: Hsu, Mei-Yin
; APPLICANT: Eagle, Susan
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: PROKARYOTIC REVERSE TRANSCRIPTASE
```

```
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,031A
; FILING DATE: 03-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377(913).5888P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-808-031A-29

Query Match      80.5%; Score 33; DB 8; Length 546;
Best Local Similarity 77.8%; Pred. No. 3.3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 QIIEKLIK 9
      111111111
Db      511 QIIEQLIK 519

RESULT 9
US-10-205-641-1
; Sequence 1, Application US/10205641
; Publication No. US20030087379A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Assay for Identifying Inhibitors of HIV
; TITLE OF INVENTION: RT Dimerization
; FILE REFERENCE: 13/085
; CURRENT APPLICATION NUMBER: US/10/205,641
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,883
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 560
; TYPE: PRT
; ORGANISM: HXB2 HIV-1 P66C2P
US-10-205-641-1

Query Match      80.5%; Score 33; DB 15; Length 560;
Best Local Similarity 77.8%; Pred. No. 3.4e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 QIIEKLIK 9
      111111111
Db      520 QIIEQLIK 528

RESULT 10
```

```

US-10-059-271-87
; Sequence 87, Application US/10059271
; Publication No. US20030082208A1
; GENERAL INFORMATION:
; APPLICANT: REPKE, HEINRICH
; APPLICANT: BUDEE, ECKHARD
; APPLICANT: NICOLAUS, STEFAN
; TITLE OF INVENTION: PROTEIN HAVING MULTIPLE ANTIGEN/EPITOPE SEQUENCES AND
; FILE REFERENCE: ALBRE-22
; CURRENT APPLICATION NUMBER: US/10/059,271
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: DE 101 06 295
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-059-271-87

Query Match      80.5%; Score 33; DB 15; Length 561;
Best Local Similarity 77.8%; Pred. No. 3.4e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 521 QIIEQLIKK 529

RESULT 11
US-09-735-487-14
; Sequence 14, Application US/09735487
; Patent No. US20020042679A1
; GENERAL INFORMATION:
; APPLICANT: de BETHUNE, Marie-Pierre
; APPLICANT: HERTOES, Kurt
; APPLICANT: PAUWELS, Rudi
; TITLE OF INVENTION: METHOD OF MANAGING THE CHEMOTHERAPY OF PATIENTS WHO ARE
; TITLE OF INVENTION: HIV POSITIVE BASED ON THE PHENOTYPIC DRUG SENSITIVITY
; FILE REFERENCE: 1377-125P
; CURRENT APPLICATION NUMBER: US/09/735,487
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 09/117,217
; PRIOR FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 562
; TYPE: PRT
; ORGANISM: HIV-HXB2
US-09-735-487-14

Query Match      80.5%; Score 33; DB 9; Length 562;
Best Local Similarity 77.8%; Pred. No. 3.4e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 520 QIIEQLIKK 528

RESULT 12
US-09-952-060-2
; Sequence 2, Application US/09952060
; Publication No. US20030044421A1
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.

```

```

; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 850
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized HIV-1 pol
US-09-952-060-2

Query Match      80.5%; Score 33; DB 11; Length 850;
Best Local Similarity 77.8%; Pred. No. 5.2e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 522 QIIEQLIKK 530

RESULT 13
US-09-952-060-4
; Sequence 4, Application US/09952060
; Publication No. US20030044421A1
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 850
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```
; FEATURE:
; OTHER INFORMATION: Codon optimized DNA inactivated Pol (IA-Pol)
US-09-952-060-4

Query Match      80.5%; Score 33; DB 11; Length 850;
Best Local Similarity 77.8%; Pred. No. 5.2e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9
    1111:11:1
Db 522 QIIEQLIK 530

RESULT 14
US-10-059-271-88
; Sequence 88, Application US/10059271
; Publication No. US20030082208A1
; GENERAL INFORMATION:
; APPLICANT: REPKE, HEINRICH
; APPLICANT: BUDE, ECKHARD
; APPLICANT: NICOLAUS, STEFAN
; TITLE OF INVENTION: PROTEIN HAVING MULTIPLE ANTIGEN/EPITOPE SEQUENCES AND
; TITLE OF INVENTION: BEING IMMOBILIZED
; FILE REFERENCE: ALBRE-22
; CURRENT APPLICATION NUMBER: US/10/059,271
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: DE 101 06 295
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 88
; LENGTH: 851
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (65)..(66)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (68)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (71)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (74)..(75)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (80)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (83)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (86)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (90)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (93)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (96)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (101)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (103)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (105)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (109)
; OTHER INFORMATION: Asp or Glu
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (113)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (115)..(116)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (118)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (290)..(291)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (293)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (296)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (299)..(300)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (305)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (308)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (311)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (315)
; OTHER INFORMATION: Any amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (318)
; OTHER INFORMATION: Any amino acid
; FEATURE:
```

```
NAME/KEY: MOD_RES
LOCATION: (321)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (324)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (326)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (328)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (330)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (334)
OTHER INFORMATION: Asp or Glu
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (338)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (340)...(341)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (343)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (404)...(405)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (407)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (410)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (413)...(414)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (419)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (422)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (425)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (429)
OTHER INFORMATION: Any amino acid
FEATURE:
NAME/KEY: MOD_RES
```

Query Match 80.5%; Score 33; DB 15; Length 851;  
Best Local Similarity 77.8%; Pred. No. 5.2e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
||||:|:|  
Db 753 QIIEQLIKK 761

## RESULT 15

```
US-09-952-060-6
; Sequence 6, Application US/09952060
; Publication No. US20030044421A1
; GENERAL INFORMATION:
; APPLICANT: Emini, Emilio A.
; APPLICANT: Youil, Rima
; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized HIV-1 pol fused to an amino
; OTHER INFORMATION: terminal localized leader sequence
US-09-952-060-6
```

Query Match 80.5%; Score 33; DB 11; Length 875;  
Best Local Similarity 77.8%; Pred. No. 5.3e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
||||:|:|  
Db 547 QIIEQLIKK 555

Search completed: July 29, 2003, 18:32:05  
Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 18:21:11 ; Search time 374 seconds  
(without alignments)  
20.944 Million cell updates/sec

Title: US-09-994-617-4

Perfect score: 41

Sequence: 1 QIIEKLIK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 5580241 seqs, 870357830 residues

Total number of hits satisfying chosen parameters: 5580241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main :

1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097A\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US097B\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US099A\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US099B\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
28: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
29: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*  
30: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep.\*  
31: /cgn2\_6/ptodata/1/paa/US160\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	41	100.0	9 25	US-09-994-617-4
2	41	100.0	958 22	US-09-791-537-148976

3	41	100.0	1002	22	US-09-791-537-101358	Sequence 101358, A
4	41	100.0	1003	22	US-09-791-537-35906	Sequence 35906, A
5	41	100.0	1003	22	US-09-791-537-35929	Sequence 35929, A
6	41	100.0	1003	22	US-09-791-537-71374	Sequence 71374, A
7	41	100.0	1003	22	US-09-791-537-151927	Sequence 151927, A
8	41	100.0	1433	22	US-09-791-537-77194	Sequence 77194, A
9	41	100.0	1433	22	US-09-791-537-127466	Sequence 127466, A
10	38	92.7	1001	22	US-09-791-537-84791	Sequence 84791, A
11	38	92.7	1002	22	US-09-791-537-123216	Sequence 123216, A
12	37	90.2	421	22	US-09-791-537-75083	Sequence 75083, A
13	37	90.2	460	22	US-09-791-537-141719	Sequence 141719, A
14	37	90.2	493	30	US-10-431-652-6753	Sequence 6753, Ap
15	37	90.2	1003	22	US-09-791-537-78701	Sequence 78701, A
16	37	90.2	1003	22	US-09-791-537-103163	Sequence 103163, A
17	37	90.2	1003	22	US-09-791-537-140829	Sequence 140829, A
18	36	87.8	343	30	US-10-437-963-135873	Sequence 135873, A
19	36	87.8	564	22	US-09-791-537-35530	Sequence 35530, A
20	36	87.8	780	28	US-10-282-122A-47011	Sequence 47011, A
21	36	87.8	999	22	US-09-791-537-71397	Sequence 71397, A
22	36	87.8	1805	6	US-08-241-640-2	Sequence 2, Appl1
23	36	87.8	1805	28	US-10-219-051B-14398	Sequence 14398, A
24	35	85.4	49	1	PCT-US00-26324B-5660	Sequence 5660, Ap
25	35	85.4	49	27	US-10-106-698-5670	Sequence 5670, Ap
26	35	85.4	49	28	US-10-221-279-10300	Sequence 10300, A
27	35	85.4	97	21	US-09-724-676-86924	Sequence 86924, A
28	35	85.4	97	21	US-09-724-676A-86924	Sequence 86924, A
29	35	85.4	113	21	US-09-724-676-86918	Sequence 86918, A
30	35	85.4	113	21	US-09-724-676A-86918	Sequence 86918, A
31	35	85.4	123	21	US-09-724-676-86928	Sequence 86928, A
32	35	85.4	123	21	US-09-724-676A-86928	Sequence 86928, A
33	35	85.4	128	21	US-09-724-676-86923	Sequence 86923, A
34	35	85.4	128	21	US-09-724-676-86923	Sequence 86923, A
35	35	85.4	128	21	US-09-724-676A-86922	Sequence 86922, A
36	35	85.4	128	21	US-09-724-676A-86923	Sequence 86923, A
37	35	85.4	139	21	US-09-724-676-86921	Sequence 86921, A
38	35	85.4	139	21	US-09-724-676A-86921	Sequence 86921, A
39	35	85.4	144	21	US-09-724-676-86916	Sequence 86916, A
40	35	85.4	144	21	US-09-724-676-86917	Sequence 86917, A
41	35	85.4	144	21	US-09-724-676A-86916	Sequence 86916, A
42	35	85.4	144	21	US-09-724-676A-86917	Sequence 86917, A
43	35	85.4	154	21	US-09-724-676-86926	Sequence 86926, A
44	35	85.4	154	21	US-09-724-676-86927	Sequence 86927, A
45	35	85.4	154	21	US-09-724-676A-86926	Sequence 86926, A

#### ALIGNMENTS

RESULT 1  
US-09-994-617-4  
; Sequence 4, Application US/09994617  
; GENERAL INFORMATION:  
; APPLICANT: TAKIGUCHI, Masafumi  
; APPLICANT: MIWA, Kiyoshi  
; TITLE OF INVENTION: HIV-SPECIFIC CTL INDUCING PEPTIDES AND MEDICAMENTS FOR PREVENT  
; FILE REFERENCE: 216457USOCNT  
; CURRENT APPLICATION NUMBER: US/09/994,617  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/03455  
; PRIOR FILING DATE: 2000-05-29  
; PRIOR APPLICATION NUMBER: JP11-149589  
; PRIOR FILING DATE: 1999-05-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-994-617-4

Query Match 100.0%; Score 41; DB 25; Length 9;  
Best Local Similarity 100.0%; Pred. No. 5.2e+06;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 1 QIIEKLIK 9

## RESULT 2

US-09-791-537-148976  
; Sequence 148976, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 148976  
; LENGTH: 958  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-148976

Query Match 100.0%; Score 41; DB 22; Length 958;

Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 669 QIIEKLIK 677

## RESULT 3

US-09-791-537-101358  
; Sequence 101358, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 101358  
; LENGTH: 1002  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-101358

Query Match 100.0%; Score 41; DB 22; Length 1002;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 674 QIIEKLIK 682

## RESULT 4

US-09-791-537-35906  
; Sequence 35906, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB

; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 35906  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-35906

Query Match 100.0%; Score 41; DB 22; Length 1003;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 675 QIIEKLIK 683

## RESULT 5

US-09-791-537-35929  
; Sequence 35929, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 35929  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-35929

Query Match 100.0%; Score 41; DB 22; Length 1003;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 675 QIIEKLIK 683

## RESULT 6

US-09-791-537-71374  
; Sequence 71374, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 71374  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-71374

Query Match 100.0%; Score 41; DB 22; Length 1003;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 675 QIIEKLIK 683

## RESULT 7

US-09-791-537-151927  
; Sequence 151927, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN MEMBERS OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 151927  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (732)..(732)  
; OTHER INFORMATION: X is an unknown amino acid  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (931)..(931)  
; OTHER INFORMATION: X is an unknown amino acid  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (936)..(936)  
; OTHER INFORMATION: X is an unknown amino acid  
US-09-791-537-151927

Query Match 100.0%; Score 41; DB 22; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 675 QIIEKLIK 683

## RESULT 8

US-09-791-537-77194  
; Sequence 77194, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 77194  
; LENGTH: 1433  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-77194

Query Match 100.0%; Score 41; DB 22; Length 1433;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |

Db 1105 QIIEKLIK 1113

## RESULT 9

US-09-791-537-127466  
; Sequence 127466, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 127466  
; LENGTH: 1433  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-127466

Query Match 100.0%; Score 41; DB 22; Length 1433;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 1105 QIIEKLIK 1113

## RESULT 10

US-09-791-537-84791  
; Sequence 84791, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 84791  
; LENGTH: 1001  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-84791

Query Match 92.7%; Score 38; DB 22; Length 1001;  
Best Local Similarity 88.9%; Pred. No. 8.7e+02;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
Db 673 QIIEKLIK 681

## RESULT 11

US-09-791-537-123216  
; Sequence 123216, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 123216  
; LENGTH: 1002  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-123216

Query Match 92.7%; Score 38; DB 22; Length 1002;  
Best Local Similarity 88.9%; Pred. No. 8.7e+02;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
DB 674 QIIEKLIQK 682

## RESULT 12

US-09-791-537-75083  
; Sequence 75083, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 75083  
; LENGTH: 421  
; TYPE: PRT  
; ORGANISM: Archaeoglobus fulgidus  
US-09-791-537-75083

Query Match 90.2%; Score 37; DB 22; Length 421;  
Best Local Similarity 77.8%; Pred. No. 4.7e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
DB 407 QIIEKLIK 415

## RESULT 13

US-09-791-537-141719  
; Sequence 141719, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 141719  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
US-09-791-537-141719

Query Match 90.2%; Score 37; DB 22; Length 460;  
Best Local Similarity 77.8%; Pred. No. 5.2e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
DB 122 EIEKLVK 130

## RESULT 14

US-10-431-652-6753  
; Sequence 6753, Application US/10431652  
; GENERAL INFORMATION:  
; APPLICANT: Breton, Gary L.  
; APPLICANT: Bush, David  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: PATH03-08  
; CURRENT APPLICATION NUMBER: US/10/431,652  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: US 09/328,352  
; PRIOR FILING DATE: 1999-06-04  
; PRIOR APPLICATION NUMBER: US 60/088,701  
; PRIOR FILING DATE: 1998-06-09  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6753  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-10-431-652-6753

Query Match 90.2%; Score 37; DB 30; Length 493;  
Best Local Similarity 88.9%; Pred. No. 5.6e+02;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
DB 310 QIIEKLIK 318

## RESULT 15

US-09-791-537-78701  
; Sequence 78701, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 78701  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-09-791-537-78701

Query Match 90.2%; Score 37; DB 22; Length 1003;  
Best Local Similarity 88.9%; Pred. No. 1.3e+03;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
| | | | | | | |  
DB 675 QIIEKLIK 683

Search completed: July 29, 2003, 18:30:48  
Job time : 375 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 18:21:31 ; Search time 13 seconds  
(without alignments)  
2.843 Million cell updates/sec

Title: US-09-994-617-4

Perfect score: 41

Sequence: 1 QIEKLEK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 41799 seqs, 4106219 residues

Total number of hits satisfying chosen parameters: 41799

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	78.0	34	1 PCT-US03-18890-5	Sequence 5, Appli
2	32	78.0	34	1 PCT-US03-18890-9	Sequence 9, Appli
3	32	78.0	34	1 PCT-US03-18890-12	Sequence 12, Appli
4	31	75.6	832	5 US-09-784-553C-2	Sequence 2, Appli
5	31	75.6	832	6 US-10-209-201C-2	Sequence 2, Appli
6	29	70.7	73	6 US-10-375-209A-16	Sequence 16, Appli
7	29	70.7	224	6 US-10-273-573-6999	Sequence 6999, Ap
8	29	70.7	224	6 US-10-273-573-8802	Sequence 8802, Ap
9	29	70.7	416	5 US-09-291-417D-5	Sequence 5, Appli
10	29	70.7	500	6 US-10-014-099F-57	Sequence 57, Appli
11	28	68.3	73	6 US-10-375-209A-27	Sequence 27, Appli
12	28	68.3	102	6 US-10-273-573-6409	Sequence 6409, Ap
13	28	68.3	247	1 PCT-US02-18638A-236	Sequence 236, App
14	28	68.3	247	6 US-10-273-573-9785	Sequence 9785, Ap
15	28	68.3	261	6 US-10-411-910A-175	Sequence 175, App
16	28	68.3	647	6 US-10-275-595A-8	Sequence 8, Appli
17	28	68.3	1086	6 US-10-275-595A-9	Sequence 9, Appli
18	28	68.3	1805	5 US-09-820-843B-73	Sequence 73, Appli
19	27	65.9	34	1 PCT-US03-18890-10	Sequence 10, Appli
20	27	65.9	86	6 US-10-463-190-134	Sequence 134, App
21	27	65.9	132	6 US-10-273-573-6541	Sequence 6541, Ap
22	27	65.9	161	6 US-10-287-971-32	Sequence 32, Appli
23	27	65.9	180	6 US-10-463-190-44	Sequence 44, Appli
24	27	65.9	180	6 US-10-287-971-34	Sequence 34, Appli
25	27	65.9	193	6 US-10-273-573-9084	Sequence 9084, Ap
26	27	65.9	209	6 US-10-370-481-19	Sequence 19, Appli

Sequence 18, Appli  
Sequence 55, Appli  
Sequence 2, Appli  
Sequence 6908, Ap  
Sequence 20, Appli  
Sequence 6544, Ap  
Sequence 10, Appli  
Sequence 30, Appli  
Sequence 2, Appli  
Sequence 7, Appli  
Sequence 101, App  
Sequence 295, App  
Sequence 3325, Ap  
Sequence 727, App  
Sequence 4, Appli  
Sequence 1, Appli  
Sequence 10663, A  
Sequence 10899, A  
Sequence 22, Appli

27 27 65.9 214 6 US-10-370-481-18  
28 27 65.9 240 6 US-10-411-910A-55  
29 27 65.9 250 6 US-10-370-480-2  
30 27 65.9 297 6 US-10-273-573-6908  
31 27 65.9 317 6 US-10-220-475A-20  
32 27 65.9 348 6 US-10-273-573-6544  
33 27 65.9 364 6 US-10-451-901-10  
34 27 65.9 368 6 US-10-370-481-30  
35 27 65.9 563 1 PCT-US03-19660-2  
36 27 65.9 701 6 US-10-333-120A-7  
37 27 65.9 788 1 PCT-US02-18638A-101  
38 27 65.9 975 6 US-10-294-433-295  
39 27 65.9 1147 7 US-60-478-196-3325  
40 27 65.9 1186 6 US-10-294-433-727  
41 27 65.9 2463 1 PCT-US03-20841-4  
42 27 65.9 2548 5 US-09-851-682B-1  
43 27 65.9 3262 6 US-10-273-573-10663  
44 27 65.9 3262 6 US-10-273-573-10899  
45 27 65.9 4563 6 US-10-370-480-22

#### ALIGNMENTS

##### RESULT 1

PCT-US03-18890-5

; Sequence 5, Application PC/TUS0318890  
; GENERAL INFORMATION:

; APPLICANT: Beth Israel Deaconess Medical Center, Inc.

; APPLICANT: Chorem, Michael

; APPLICANT: Rosenblatt, Michael

; TITLE OF INVENTION: ANALOGS OF PARATHYROID HORMONE AND PTH-RELATED PROTEIN AS BONE

; TITLE OF INVENTION: AGENTS

; FILE REFERENCE: B00662.70054.WO

; CURRENT APPLICATION NUMBER: PCT/US03/18890

; CURRENT FILING DATE: 2003-06-13

; PRIOR APPLICATION NUMBER: 60/388,918

; PRIOR FILING DATE: 2002-06-13

; PRIOR APPLICATION NUMBER: 60/398,005

; PRIOR FILING DATE: 2002-07-23

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: Patent version 3.2

; SEQ ID NO 5

; LENGTH: 34

; TYPE: PRT

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: Designed sequence

PCT-US03-18890-5

Query Match 78.0%; Score 32; DB 1; Length 34;

Best Local Similarity 55.6%; Pred. No. 2.1;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIEKLEK 9

Db 22 ELLEKLEK 30

##### RESULT 2

PCT-US03-18890-9

; Sequence 9, Application PC/TUS0318890

; GENERAL INFORMATION:

; APPLICANT: Beth Israel Deaconess Medical Center, Inc.

; APPLICANT: Chorem, Michael

; APPLICANT: Rosenblatt, Michael

; TITLE OF INVENTION: ANALOGS OF PARATHYROID HORMONE AND PTH-RELATED PROTEIN AS BONE

; TITLE OF INVENTION: AGENTS

; FILE REFERENCE: B00662.70054.WO

; CURRENT APPLICATION NUMBER: PCT/US03/18890

; CURRENT FILING DATE: 2003-06-13

; PRIOR APPLICATION NUMBER: 60/388,918

; PRIOR FILING DATE: 2002-06-13

; PRIOR APPLICATION NUMBER: 60/398,005  
; PRIOR FILING DATE: 2002-07-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 9  
; LENGTH: 34  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Designed sequence  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (32)..(32)  
; OTHER INFORMATION: Xaa - 2-thienyl-alanine (Thi)  
PCT-US03-18890-9

Query Match 78.0%; Score 32; DB 1; Length 34;  
Best Local Similarity 55.6%; Pred. No. 2.1;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 22 ELLEKLIK 30

## RESULT 3

PCT-US03-18890-12  
; Sequence 12, Application PC/TUS0318890  
; GENERAL INFORMATION:  
; APPLICANT: Beth Israel Deaconess Medical Center, Inc.  
; APPLICANT: Chorev, Michael  
; APPLICANT: Rosenblatt, Michael  
; TITLE OF INVENTION: ANALOGS OF PARATHYROID HORMONE AND PTH-RELATED PROTEIN AS BONE AN  
; FILE REFERENCE: B00662.70054 WO  
; CURRENT APPLICATION NUMBER: PCT/US03/18890  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: 60/388,918  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: 60/398,005  
; PRIOR FILING DATE: 2002-07-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12  
; LENGTH: 34  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Designed sequence  
PCT-US03-18890-12

Query Match 78.0%; Score 32; DB 1; Length 34;  
Best Local Similarity 66.7%; Pred. No. 2.1;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 22 EFLEKLIK 30

## RESULT 4

US-09-784-553C-2  
; Sequence 2, Application US/09784553C  
; GENERAL INFORMATION:  
; APPLICANT: ZHOU, MING-MING  
; APPLICANT: AGGARWAL, ANEEL  
; TITLE OF INVENTION: METHODS OF IDENTIFYING MODULATORS OF BROMODOMAINS  
; FILE REFERENCE: 2459-1-003 CIP  
; CURRENT APPLICATION NUMBER: US/09/784,553C  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: 09/510,314  
; PRIOR FILING DATE: 2000-02-22  
; NUMBER OF SEQ ID NOS: 63

; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 832  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-784-553C-2

Query Match 75.6%; Score 31; DB 5; Length 832;  
Best Local Similarity 66.7%; Pred. No. 88;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 668 EIIKKLIER 676

## RESULT 5

US-10-209-201C-2  
; Sequence 2, Application US/10209201C  
; GENERAL INFORMATION:  
; APPLICANT: Verdin, Eric  
; APPLICANT: Bruland, Joan  
; APPLICANT: Ott, Melanie  
; APPLICANT: Zhou, Ming-Ming  
; APPLICANT: Aggarwal, Aneel  
; TITLE OF INVENTION: Methods of Identifying Modulators of Bromodomains  
; FILE REFERENCE: 2459-1-003CIPDIV  
; CURRENT APPLICATION NUMBER: US/10/209,201C  
; CURRENT FILING DATE: 2002-07-31  
; PRIOR APPLICATION NUMBER: 09/784,553  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: 09/510,314  
; PRIOR FILING DATE: 2000-02-22  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 832  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-209-201C-2

Query Match 75.6%; Score 31; DB 6; Length 832;  
Best Local Similarity 66.7%; Pred. No. 88;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9  
Db 668 EIIKKLIER 676

## RESULT 6

US-10-375-209A-16  
; Sequence 16, Application US/10375209A  
; GENERAL INFORMATION:  
; APPLICANT: McDonald, John R.  
; APPLICANT: Coggin, Philip  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING SECONDARY TISSUE DAMAGE  
; FILE REFERENCE: 25020-601E  
; CURRENT APPLICATION NUMBER: US/10/375,209A  
; CURRENT FILING DATE: 2003-02-24  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: homo sapien  
; FEATURE:  
; OTHER INFORMATION: Human Chemokine Polypeptide: MGSA/GRO-Alpha  
US-10-375-209A-16

Query Match 70.7%; Score 29; DB 6; Length 73;  
Best Local Similarity 62.5%; Pred. No. 16;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 2 IIEKLEIK 9  
I:::||||  
Db 58 IVKLEIK 65

RESULT 7  
US-10-273-573-6999  
; Sequence 6999, Application US/10273573  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 21272-066  
; CURRENT APPLICATION NUMBER: US/10/273,573  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 09/522,929  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: 09/770,160  
; PRIOR FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 10994  
; SOFTWARE: Custom  
; SEQ ID NO 6999  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-273-573-6999

Query Match 70.7%; Score 29; DB 6; Length 224;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QIEKLEIK 9  
I:::||||  
Db 190 QLMKLEIK 198

RESULT 8  
US-10-273-573-8802  
; Sequence 8802, Application US/10273573  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 21272-066  
; CURRENT APPLICATION NUMBER: US/10/273,573  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 09/522,929  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: 09/770,160  
; PRIOR FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 10994  
; SOFTWARE: Custom  
; SEQ ID NO 8802  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-273-573-8802

Query Match 70.7%; Score 29; DB 6; Length 224;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QIEKLEIK 9  
I:::||||  
Db 190 QLMKLEIK 198

RESULT 9  
US-09-291-417D-5  
; Sequence 5, Application US/09291417D  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: MARTINEZ, RICARDO

; APPLICANT: WHYTE, DAVID  
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
; FILE REFERENCE: 038602/0329  
; CURRENT APPLICATION NUMBER: US/09/291,417D  
; CURRENT FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: 60/081,784  
; PRIOR FILING DATE: 1998-04-14  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 416  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-291-417D-5

Query Match 70.7%; Score 29; DB 5; Length 416;  
Best Local Similarity 55.6%; Pred. No. 96;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QIEKLEIK 9  
:::|||||  
Db 398 KMVKLEIK 406

RESULT 10  
US-10-014-099F-57  
; Sequence 57, Application US/10014099F  
; GENERAL INFORMATION:  
; APPLICANT: KOEHN, Ralf  
; APPLICANT: FELDER, Susanne  
; APPLICANT: SCHWENK, Frieder  
; APPLICANT: KUETER-LUKS, Birgit  
; APPLICANT: FAUST, Nicole  
; TITLE OF INVENTION: Modified Recombinase  
; FILE REFERENCE: 012787wo/JH/ml  
; CURRENT APPLICATION NUMBER: US/10/014,099F  
; CURRENT FILING DATE: 2001-11-12  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 57  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Cisa recombinase  
US-10-014-099F-57

Query Match 70.7%; Score 29; DB 6; Length 500;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QIEKLEIK 9  
:::||||  
Db 385 ELIEKLEIK 393

RESULT 11  
US-10-375-209A-27  
; Sequence 27, Application US/10375209A  
; GENERAL INFORMATION:  
; APPLICANT: McDONALD, John R.  
; APPLICANT: COGGINS, Philip  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING SECONDARY TISSUE DAMAGE  
; FILE REFERENCE: 25020-601E  
; CURRENT APPLICATION NUMBER: US/10/375,209A  
; CURRENT FILING DATE: 2003-02-24  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: homo sapien  
; FEATURE:  
; OTHER INFORMATION: Human Chemokine Polypeptide: MGSA/GRO-gamma

## US-10-375-209A-27

Query Match 68.3%; Score 28; DB 6; Length 73;  
Best Local Similarity 55.6%; Pred. No. 24;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
Db 61 KIIEKLINK 69

## RESULT 12

US-10-273-573-6409  
; Sequence 6409, Application US/10273573  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 21272-066  
; CURRENT APPLICATION NUMBER: US/10/273,573  
; PRIOR FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 09/522,929  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: 09/770,160  
; PRIOR FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 10994  
; SOFTWARE: Custom  
; SEQ ID NO 6409  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(102)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-273-573-6409

Query Match 68.3%; Score 28; DB 6; Length 102;  
Best Local Similarity 66.7%; Pred. No. 34;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
Db 64 QIIEVLLEK 72

## RESULT 13

PCT-US02-18638A-236  
; Sequence 236, Application PC/TUS0218638A  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc. et al.  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035PC  
; CURRENT APPLICATION NUMBER: PCT/US02/18638A  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 236  
; LENGTH: 247  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US02-18638A-236

Query Match 68.3%; Score 28; DB 1; Length 247;  
Best Local Similarity 62.5%; Pred. No. 84;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 8  
Db 46 QILKKIIE 53

## RESULT 14

US-10-273-573-9785  
; Sequence 9785, Application US/10273573  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 21272-066  
; CURRENT APPLICATION NUMBER: US/10/273,573  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 09/522,929  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: 09/770,160  
; PRIOR FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 10994  
; SOFTWARE: Custom  
; SEQ ID NO 9785  
; LENGTH: 247  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (14)...(216)  
; OTHER INFORMATION: ATP synthase subunit D domain identified by Pfam, accession  
; US-10-273-573-9785

Query Match 68.3%; Score 28; DB 6; Length 247;  
Best Local Similarity 62.5%; Pred. No. 84;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 8  
Db 46 QILKKIIE 53

## RESULT 15

US-10-411-910A-175  
; Sequence 175, Application US/10411910A  
; GENERAL INFORMATION:  
; APPLICANT: Dillion, Harrison F.  
; TITLE OF INVENTION: Methods and Compositions for Evolving Hydrogenase Genes  
; FILE REFERENCE: H2041203-P  
; CURRENT APPLICATION NUMBER: US/10/411,910A  
; CURRENT FILING DATE: 2003-04-12  
; NUMBER OF SEQ ID NOS: 343  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 175  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
US-10-411-910A-175

Query Match 68.3%; Score 28; DB 6; Length 261;  
Best Local Similarity 57.1%; Pred. No. 89;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEKLIK 9  
Db 248 VEKMEK 254

Search completed: July 29, 2003, 18:31:08  
Job time : 14 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 18:17:55 ; Search time 39 Seconds  
(without alignments)  
22.193 Million cell updates/sec

Title: US-09-994-617-4

Perfect score: 41

Sequence: 1 QIEKLIK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	95.1	1616	2 T16600	vitellogenin vit-1
2	37	90.2	421	2 H69490	formylmethanofuran
3	37	90.2	460	2 E72341	cysteine-tRNA liga
4	37	90.2	651	2 S24609	cytoskeletal prote
5	36	87.8	366	2 H87332	articulin, probabl
6	36	87.8	375	2 T28710	hypothetical prote
7	36	87.8	780	2 B70112	DNA mismatch repai
8	36	87.8	1805	2 A34736	nestin - rat
9	35	85.4	397	2 A35136	cellulase (EC 3.2.
10	35	85.4	415	1 K1FFPG	phosphoglycerate k
11	35	85.4	432	2 G90268	conserved hypothet
12	35	85.4	442	2 C84985	trigger factor [im
13	35	85.4	1105	2 S76557	carbamoyl-phosphat
14	34	82.9	136	2 T09583	acyl carrier prote
15	34	82.9	208	2 E90201	conserved hypothet
16	34	82.9	287	2 T32029	hypothetical prote
17	34	82.9	288	2 B72272	iron-sulfur cluste
18	34	82.9	332	1 DEHULC	L-lactate dehydrog
19	34	82.9	389	2 D82223	ribonucleoside-dip
20	34	82.9	389	2 T25884	hypothetical prote
21	34	82.9	437	2 A69075	formylmethanofuran
22	34	82.9	579	2 A64100	inner membrane cop
23	34	82.9	644	2 T37692	probable transport
24	34	82.9	902	2 T01668	pol polyprotein -
25	34	82.9	1108	2 E71104	probable cell divi
26	34	82.9	1804	2 T34518	nestin - golden ha
27	33	80.5	112	2 T47738	hypothetical prote
28	33	80.5	138	2 S11523	ribonuclease H (EC
29	33	80.5	150	2 E71870	hypothetical prote

30 33 80.5 265 2 D84643 probable ATP-depen  
31 33 80.5 345 2 S04280 psba intron 1 prot  
32 33 80.5 347 2 G64321 cell division prot  
33 33 80.5 356 2 G95897 hypothetical prote  
34 33 80.5 377 2 D81841 ribonucleoside-dip  
35 33 80.5 384 2 G81100 ribonucleoside-dip  
36 33 80.5 487 1 A64472 carbamoyl-phosphat  
37 33 80.5 495 2 H70391 cysteine-tRNA liga  
38 33 80.5 559 2 A47175 reverse transcript  
39 33 80.5 559 2 B47175 reverse transcript  
40 33 80.5 564 2 A89956 hypothetical prote  
41 33 80.5 610 2 E72358 cell division prot  
42 33 80.5 674 2 S52240 DNA ligase (NAD) (  
43 33 80.5 1002 1 GNLJND HIV-1 retropepsin  
44 33 80.5 1002 2 S54378 pol polyprotein -  
45 33 80.5 1003 1 GNVVLV HIV-1 retropepsin

#### ALIGNMENTS

##### RESULT 1

T16600

vitellogenin vit-1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000

C:Accession: T16600

R:Bentley, D.

submitted to the EMBL Data Library, September 1995

A:Description: The sequence of C. elegans cosmid K09F5.

A:Reference number: Z18544

A:Accession: T16600

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1616 <BEN>

A:Cross-references: EMBL:U37430; NID:gl945496; PID:gl945496; PIDN:AAB52675.1; GSPDB:G

A:Experimental source: strain Bristol N2; clone K09F5

C:Genetics:

A:Gene: CESP: vit-1

A:Map position: X

A:introns: 291/3; 386/3; 1448/3; 1531/1

C:Superfamily: vitellogenin

Query Match 95.1%; Score 39; DB 2; Length 1616;  
Best Local Similarity 88.9%; Pred. No. 58;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIEKLIK 9

I:|||||

Db 791 QIEKLIK 799

##### RESULT 2

H69490

formylmethanofuran dehydrogenase (tungsten) chain B homolog (fwdB-2) - Archaeoglobus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999

C:Accession: H69490

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; NUID:98049343; PMID:9389475

A:Accession: H69490

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-421 <KLE>

A:Cross-references: GB:AE000970; GB:AE000782; NID:g2689293; PIDN:AAB89326.1; PID:g264

C:Superfamily: formylmethanofuran dehydrogenase (molybdenum) beta chain

C:Keywords: iron-sulfur protein; metalloprotein; molybdenum; molybdopterin

```

Query Match      90.2%  Score 37;  DB 2;  Length 421;
Best Local Similarity 77.8%  Pred. No. 34;
Matches 7;  Conservative 2;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 1 QIIEKLEIK 9
    ||:||||:|
Db 407 QILEKLEK 415

RESULT 3
E72341
cysteine-tRNA ligase (EC 6.1.1.16) - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 03-Jun-2002
C:Accession: E72341
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: E72341
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <ARN>
A:Cross-references: GB:AE001743; GB:AE000512; NID:g4981241; PIDN:AAD35801.1; PID:g498124
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0719
C:Superfamily: cysteine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match      90.2%  Score 37;  DB 2;  Length 460;
Best Local Similarity 77.8%  Pred. No. 38;
Matches 7;  Conservative 2;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 1 QIIEKLEIK 9
    ||:||||:|
Db 122 EIIEKLEK 130

RESULT 4
S24609
cytoskeletal protein - Euglena gracilis
C:Species: Euglena gracilis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 01-Dec-2000
C:Accession: S24609; A43417
R:Bouck, G.B.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24609
A:Accession: S24609
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-651 <BOU>
A:Cross-references: EMBL:Z13962; NID:g18403; PID:g18404
R:Marrs, J.A.; Bouck, G.B.
J. Cell Biol. 118, 1465-1475, 1992
A:Title: The two major membrane skeletal proteins (articulins) of Euglena gracilis defin
A:Reference number: A43417; MUID:92394973; PMID:1522117
A:Accession: A43417
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-145,'T',147-187,'T',189-348,'T',350-651 <MAR>
A:Note: sequence extracted from NCBI backbone (NCBIP:113032)

Query Match      90.2%  Score 37;  DB 2;  Length 651;
Best Local Similarity 66.7%  Pred. No. 53;
Matches 6;  Conservative 3;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 1 QIIEKLEIK 9
    ||:||||:|
Db 479 QIIEKLEK 487

```

## RESULT 5

H87332

articulin, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: H87332

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

n. J.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: H87332

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-366 &lt;SPO&gt;

A:Cross-references: GB:AE005673; NID:g13421894; PIDN:AAK22660.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0675

Query Match

87.8% Score 36; DB 2; Length 366;

Best Local Similarity

66.7% Pred. No. 45;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9

||:||||:|

Db 74 QIIEKLEK 82

## RESULT 6

T28710

hypothetical protein T21D12.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T28710

R:Woessner, J.

submitted to the EMBL Data Library, August 1997

A:Description: The sequence of C. elegans cosmid T21D12.

A:Reference number: Z20514

A:Accession: T28710

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-375 &lt;WOE&gt;

A:Cross-references: EMBL:AF016687; PIDN:AAC48090.1; GSPDB:GN00022; CESP:T21D12.4

A:Experimental source: strain Bristol N2; clone T21D12

C:Genetics:

A:Gene: CESP:T21D12.4

A:Map position: 4

A:Introns: 56/73; 82/1; 144/3; 189/3; 286/3; 359/3

Query Match

87.8% Score 36; DB 2; Length 375;

Best Local Similarity

77.8% Pred. No. 46;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9

||:||||:|

Db 131 QIIEKLEK 139

## RESULT 7

B70112

DNA mismatch repair protein homolog - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C:Accession: B70112

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu

n. J.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: B70112

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-780 <KLE>

A:Cross-references: GB:AE001122; GB:AE000783; NID:g2687974; PIDN:AAAC66481.1; PID:g268797

A:Experimental source: strain B31

Query Match 87.8%; Score 36; DB 2; Length 780;

Best Local Similarity 77.8%; Pred. No. 97;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 520 EILEKLIK 528

RESULT 8

A34736

nestin - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-Jul-1990 #sequence\_revision 09-Oct-1992 #text\_change 05-Nov-1999

C:Accession: A34736

R:Lendahl, U.; Zimmerman, L.B.; McKay, R.D.G.

Cell 60, 585-595, 1990

A:Title: CNS stem cells express a new class of intermediate filament protein.

A:Reference number: A34736; MUID:90150286; PMID:1689217

A:Accession: A34736

A:Status: preliminary

A:Molecule type: DNA; mRNA

A:Residues: 1-1805 <LEN>

A:Cross-references: GB:M34384; NID:g205663; PIDN:AAA41685.1; PID:g205664

A:Note: the authors translated the codon CGG for residue 168 as Gly, ATC for residue 169

C:Keywords: coiled coil

Query Match

Best Local Similarity 87.8%; Score 36; DB 2; Length 1805;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 713 QILERLIK 721

RESULT 9

A35136

cellulase (EC 3.2.1.4) - Bacillus polymyxa

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus polymyxa

C:Date: 03-Aug-1990 #sequence\_revision 03-Aug-1990 #text\_change 24-Sep-1998

C:Accession: A35136

R:Baird, S.D.; Johnson, D.A.; Seligy, V.L.

J. Bacteriol. 172, 1576-1586, 1990

A:Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase

A:Reference number: A35136; MUID:90170877; PMID:2307659

A:Accession: A35136

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-397 <BAI>

A:Cross-references: GB:M33791; GB:M33840

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Superfamily: Xanthomonas campestris cellulase

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match

Best Local Similarity 85.4%; Score 35; DB 2; Length 397;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 127 QIMDKLIK 135

RESULT 10

KIFFPG

phosphoglycerate kinase (EC 2.7.2.3) - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 11-Jun-1999

C:Accession: S30111; S25673

R:Roselli-Rehfuess, L.; Ye, F.; Lissemore, J.L.; Sullivan, D.T.

Mol. Gen. Genet. 235, 213-220, 1992

A:Title: Structure and expression of the phosphoglycerate kinase (Pgk) gene of Drosop

A:Reference number: S30111; MUID:93101123; PMID:1465095

A:Accession: S30111

A:Molecule type: DNA

A:Residues: 1-415 <ROS>

A:Cross-references: EMBL:Z14029; NID:g11175; PIDN:CAA78404.1; PID:g11176

A:Experimental source: strain Oregon-R

A:Note: the authors translated the codon AAG for residue 139 as Cys

C:Genetics:

A:Gene: pgk

A:Cross-references: FlyBase:FBgn0003075

A:Map position: 11,23,Al-2

A:Introns: 21/2; 336/2

C:Superfamily: phosphoglycerate kinase

C:Keywords: ATP; gluconeogenesis; glycolysis; phosphotransferase

F:218,342/Binding site: ATP (Lys, Glu) #status predicted

Query Match

Best Local Similarity 85.4%; Score 35; DB 1; Length 415;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 262 KIVEKLEK 270

RESULT 11

G90268

conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C:Accession: G90268

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

arrett, R.A.; Ragap, M.A.; Senses, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90268

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-432 <KUR>

A:Cross-references: GB:AE006641; NID:gl3814346; PIDN:AAK41406.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS01155

Query Match

Best Local Similarity 85.4%; Score 35; DB 2; Length 432;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLIK 9

Db 58 QVVKLIK 66

RESULT 12

C84985

trigger factor [imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C:Accession: C84985

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A  
A:Reference number: A84930; MUID:20445173; PMID:10993077  
A:Accession: C84985  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <STO>  
A:Cross-references: GB:AF000398; GSPDB:GN00144  
A:Experimental source: strain APS  
C:Genetics:  
A:Gene: tig: BU474

Query Match 85.4%; Score 35; DB 2; Length 442;  
Best Local Similarity 66.7%; Pred. No. 83;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
|||:|:|:|  
Db 290 QIIQKIVEK 298

RESULT 13  
S76357  
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain [similarit  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Sep-2000  
C:Accession: S76357  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
S.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S76357  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1105 <KAN>  
A:Cross-references: EMBL:D64002; GB:AB001339; NID:gl001612; PIDN:BAA10403.1; PID:gl00166  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Start codon: GTG  
C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin  
C:Keywords: ligase  
F:33-1086/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homol  
F:33-494/Domain: biotin carboxylase homology <BCI>  
F:584-1024/Domain: biotin carboxylase homology <BC2>

Query Match 85.4%; Score 35; DB 2; Length 1105;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
||:|:|:|  
Db 96 EIVEKIEK 104

RESULT 14  
T09583  
acyl carrier protein - swamp oak  
C:Species: Casuarina glauca (swamp oak)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T09583  
R:Bohusz, D.  
submitted to the EMBL Data Library, February 1997  
A:Description: cDNA sequence for an acyl carrier protein from actinorhizal nodules of Ca  
A:Reference number: Z16750  
A:Accession: T09583  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-136 <BOG>  
A:Cross-references: EMBL:Y10994  
C:Superfamily: acyl carrier protein; acyl carrier protein homology  
C:Keywords: carrier protein; fatty acid biosynthesis; phosphopantetheine; phosphoprotein

F:56-127/Domain: acyl carrier protein homology <ACP>  
F:91/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 82.9%; Score 34; DB 2; Length 136;  
Best Local Similarity 87.5%; Pred. No. 38;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IIEKLEIK 9  
:|:|:|:|  
Db 126 LIEKLEIK 133

RESULT 15  
E90201  
conserved hypothetical protein [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: E90201  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder  
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: E90201  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-208 <KUR>  
A:Cross-references: GB:AE006641; NID:gl13813714; PIDN:AAK40868.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO0551

Query Match 82.9%; Score 34; DB 2; Length 208;  
Best Local Similarity 66.7%; Pred. No. 59;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
:|:|:|:|  
Db 200 EIEKLEIK 208

Search completed: July 29, 2003, 18:23:52  
Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 17:53:20 ; search time 11 seconds  
(without alignments)  
38,476 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIEKLIK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	1002	POL_HV1U4	P24740 human immun
2	39	95.1	1616	VITL_CAEEL	P55155 caenorhabdi
3	37	90.2	460	SYC_THEMA	Q9wzh8 thermotoga
4	36	87.8	343	NAS3_ORYSA	O9fxw5 oryza sativ
5	36	87.8	375	PARV_CAEEL	O16785 caenorhabdi
6	36	87.8	780	MUS2_BORBU	O51125 borrelia bu
7	36	87.8	1805	NEST_RAT	P21263 rattus norv
8	35	85.4	260	DPM1_HUMAN	O60762 homo sapien
9	35	85.4	397	GUN_PAEPO	P23548 paenibacill
10	35	85.4	415	PGK_DROME	O01604 drosophila
11	35	85.4	442	TIG_BUCAI	P57546 buchnera ap
12	35	85.4	1073	CARB_METWA	P58944 methanosarc
13	35	85.4	1081	CARB_SYNT3	Q55756 synechocyst
14	34	82.9	136	ACPI_CASGL	P93092 casuarina g
15	34	82.9	331	LDBC_HUMAN	P07864 homo sapien
16	34	82.9	579	DSBD_HAEIN	P44919 haemophilus
17	34	82.9	644	YD27_SCHPO	O13714 schizosacch
18	34	82.9	1002	POL_HV1MA	P04588 human immun
19	34	82.9	1070	CARB_METAC	O8tny4 methanosarc
20	33	80.5	138	Y4BI_RHISN	P55376 rhizobium s
21	33	80.5	345	YCX1_CHLMO	P09753 chlamydomon
22	33	80.5	347	Y174_METJA	O57638 methanococc
23	33	80.5	364	PARB_HUMAN	O9bbi1 homo sapien
24	33	80.5	365	PARB_MOUSE	O9es46 mus musculu
25	33	80.5	482	CABI_METUA	Q58773 methanococc
26	33	80.5	495	SYC_AQUAE	O67163 aquifex aeo
27	33	80.5	674	DNLJ_THESC	P49422 thermus sco
28	33	80.5	674	DNLJ_THESC	O9zfy8 thermus sp.
29	33	80.5	1002	POL_HV1EL	P04589 human immun
30	33	80.5	1002	POL_HV1ND	P18602 human immun
31	33	80.5	1002	POL_HV1RH	P05959 human immun
32	33	80.5	1002	POL_HV1Z2	P12499 human immun
33	33	80.5	1003	POL_HV1A2	P03369 human immun

34	33	80.5	1003	POL_HV1H2	P04585 human immun
35	33	80.5	1003	POL_HV1N5	P12497 human immun
36	33	80.5	1003	POL_HV1OY	P20892 human immun
37	33	80.5	1003	POL_HV1Y2	P33963 human immun
38	33	80.5	1006	POL_HV1MN	P05961 human immun
39	33	80.5	1007	POL_HV1JR	P20875 human immun
40	33	80.5	1015	POL_HV1B1	P03366 human immun
41	33	80.5	1015	POL_HV1B5	P04587 human immun
42	33	80.5	1015	POL_HV1BR	P03367 human immun
43	33	80.5	1015	POL_HV1PV	P03368 human immun
44	33	80.5	1027	POL_SIVCZ	P17283 chimpanzee
45	33	80.5	1613	VIT2_CAEEL	P05690 caenorhabdi

ALIGNMENTS

RESULT 1  
POL\_HV1U4  
ID POL\_HV1U4 STANDARD; PRT; 1002 AA.  
AC P24740;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DE 15-SEP-2003 (Rel. 42, Last annotation update)  
DE POL polyprotein [Contains: Protease (Retropepsin) (EC 3.4.23.16);  
Reverse transcriptase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)].  
GN POL.  
OS Human immunodeficiency virus type 1 (strain Ugandan / isolate U455)  
(HIV-1).  
OC Viruses; Retrovird viruses; Retroviridae; Lentiviruses.  
OX NCBI\_TaxID=11703;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91090981; PubMed=2265025;  
RA Oram J.D., Downing R.G., Roff M., Clegg J.C.S., Serwadda D.,  
Carswell J.W.;  
RT "Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic  
diversity from other HIV-1 isolates.";  
RL AIDS Res. Hum. Retroviruses 6:1073-1078(1990).  
CC -!- CATALYTIC ACTIVITY: Specific for a PI residue that is hydrophobic,  
and PI variable, but often Pro.  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
phosphomonocester.  
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
+ (DNA)(N).  
CC -!- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
DETERMINED.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
KNOWN AS THE RETROPEPSIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M62320; AAA75019.1; -;  
CC PDB; 1E27; 12-SEP-00.  
CC MEROPS; A02.001; -;  
CC InterPro; IPR001995; Aspprotease\_rtrv.  
CC InterPro; IPR001969; Aspprotease\_site.  
CC InterPro; IPR001037; Integrase\_C.  
CC InterPro; IPR003308; Integrase\_Zn.  
CC InterPro; IPR002156; RNaseH.  
CC InterPro; IPR001584; Rve.  
CC InterPro; IPR000477; RVTse.  
CC Pfam; PF00552; integrase\_1.  
CC Pfam; PF02022; integrase\_Zn; 1.  
CC Pfam; PF00075; rnaseH; 1.  
CC Pfam; PF00665; rve; 1.  
CC Pfam; PF00077; rvp; 1.

```

DR Pfam: PF00078; rvt; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;
KW Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.
FT CHAIN 56 154
FT ACT_SITE 80 80 BY SIMILARITY.
SQ SEQUENCE 1002 AA; 113595 MW; E7303380B40D97A9 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 1002;
Best Local Similarity 100.0%; Pred. NO. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 674 QIIEKLIK 682

RESULT 2
VITL_CAEEL STANDARD; PRT; 1616 AA.
AC P5155;
.DT 01-OCT-1996 (Rel. 34, Created)
.DT 01-NOV-1997 (Rel. 35, Last sequence update)
.DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 1 precursor.
GN Vit-1 OR K09F5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Bentley D.;
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Contains 1 VWF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U37430; AAB52675.1; -
CC PIR: T16600; T16600.
CC WormPep; K09F5.2; CE04746.
CC InterPro; IPR001747; Lipid_transprt_N.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF01347; Vitellogenin_N; 1.
CC Pfam; PF00094; vwd; 1.
CC SMART; SM00638; LPD_N; 1.
CC SMART; SM00216; VWD; 1.
CC Storage protein; Multigene family; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1616 VITELLOGENIN 1.
FT DOMAIN 1312 1459 VWF_D.
SQ SEQUENCE 1616 AA; 188063 MW; F173D7C452D123F5 CRC64;

Query Match 95.1%; Score 39; DB 1; Length 1616;
Best Local Similarity 88.9%; Pred. NO. 28;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 791 QIIEKLIK 799

DR Pfam: PF00078; rvt; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;
KW Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.
FT CHAIN 56 154
FT ACT_SITE 80 80 BY SIMILARITY.
SQ SEQUENCE 1002 AA; 113595 MW; E7303380B40D97A9 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 1002;
Best Local Similarity 100.0%; Pred. NO. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 674 QIIEKLIK 682

RESULT 3
ID VITL_CAEEL STANDARD; PRT; 1616 AA.
AC P5155;
.DT 01-OCT-1996 (Rel. 34, Created)
.DT 01-NOV-1997 (Rel. 35, Last sequence update)
.DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Vitellogenin 1 precursor.
GN Vit-1 OR K09F5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Bentley D.;
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Precursor of the egg-yolk proteins that are sources of
CC nutrients during embryonic development (Potential).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Contains 1 VWF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U37430; AAB52675.1; -
CC PIR: T16600; T16600.
CC WormPep; K09F5.2; CE04746.
CC InterPro; IPR001747; Lipid_transprt_N.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF01347; Vitellogenin_N; 1.
CC Pfam; PF00094; vwd; 1.
CC SMART; SM00638; LPD_N; 1.
CC SMART; SM00216; VWD; 1.
CC Storage protein; Multigene family; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1616 VITELLOGENIN 1.
FT DOMAIN 1312 1459 VWF_D.
SQ SEQUENCE 1616 AA; 188063 MW; F173D7C452D123F5 CRC64;

Query Match 95.1%; Score 39; DB 1; Length 1616;
Best Local Similarity 88.9%; Pred. NO. 28;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 791 QIIEKLIK 799

RESULT 4
ID NAS3_ORYSA STANDARD; PRT; 343 AA.
AC Q9FXW5;
.DT 28-FEB-2003 (Rel. 41, Created)
.DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

RESULT 3
ID SYNC_THEMA STANDARD; PRT; 460 AA.
AC Q9WZB8;
.DT 30-MAY-2000 (Rel. 39, Created)
.DT 30-MAY-2000 (Rel. 39, Last sequence update)
.DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cysteineyl-tRNA synthetase (EC 6.1.1.16) (Cysteine--tRNA ligase)
DE (CysRS).
DE CYS OR TM0719.
GN Thermotoga maritima.
OS Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RA MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Maik J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RA genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-cysteine + tRNA(Cys) = AMP +
CC diphosphate + L-cysteineyl-tRNA(Cys).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC Strong, to methionyl-tRNA synthetase.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001743; AAD35801.1; -
CC PIR: E72341; E72341.
CC TIGR: TM0719; -.
CC HAMAP; MF_00041; -.
CC InterPro; IPR002308; Cys_tRNA-synt_la.
CC InterPro; IPR001412; tRNA-synt_I.
CC Pfam; PF01406; tRNA-synt.le; 1.
CC PRINTS; PR00983; TRNASYNTHCYS.
CC TIGRFAMS; TIGR00435; CYS; 1.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 29 39 "HIGH" REGION.
FT SITE 264 268 "KMSK" REGION.
FT BINDING 267 267 ATP (BY SIMILARITY).
SQ SEQUENCE 460 AA; 53418 MW; 937C5A0526C4D359 CRC64;

Query Match 90.2%; Score 37; DB 1; Length 460;
Best Local Similarity 77.8%; Pred. NO. 21;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
Db 122 EIEIKLVEK 130

RESULT 4
ID NAS3_ORYSA STANDARD; PRT; 343 AA.
AC Q9FXW5;
.DT 28-FEB-2003 (Rel. 41, Created)
.DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable Nicotianamine synthase 3 (EC 2.5.1.43) (S-adenosyl-L-  
DE methionine:S-adenosyl-L-methionine:S-adenosyl-methionine 3-amino-3-  
DE carboxypropyltransferase 3) (OSNAS3).  
GN NAS3 OR OJ1165\_F02.29.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Indica-IR36; TISSUE=Root;  
RX MEDLINE=21097014; PubMed=11169192;  
RA Higuchi K., Watanabe S., Takahashi M., Kawasaki S., Nakanishi H.,  
RA Nishizawa N.-K., Mori S.;  
RT "Nicotianamine synthase gene expression differs in barley and rice  
RT under Fe-deficient conditions.";  
RL Plant J. 25:159-167(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RC Sasaki T., Matsumoto T., Yamamoto K.;  
RA "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC  
RT clone:OJ1165\_F02.";  
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Synthesizes nicotianamine, a polyamine that is the first  
CC intermediate in the synthesis of the phytylsidephores of the  
CC mugineic acid type found in gramineae which serve as a sensor for  
CC the physiological iron status within the plant, and/or might be  
CC involved in the transport of iron (By similarity).  
CC -1- CATALYTIC ACTIVITY: 3 S-adenosyl-L-methionine -> 3 5'-S-methyl-5'-  
CC thioadenosine + nicotianamine.  
CC -1- COFACTOR: Pyridoxal phosphate.  
CC -1- SIMILARITY: Contains 1 NAS domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AB023819; BAB17824.1; -;  
CC EMBL; AP003816; BAC21363.1; -;  
CC Gramene; Q9FXW5; -;  
CC InterPro: IPR004298; Nicotian\_synth.  
CC Pfam: PF03059; NAS; 1.  
CC Transferrase; Pyridoxal phosphate; Multigene family.  
FT DOMAIN 18 298  
FT NAS.  
FT CONFLICT 249 249 S -> T (IN REF. 1).  
FT CONFLICT 292 297 KPVAA -> NRPWPG (IN REF. 1).  
SQ SEQUENCE 343 AA; 36983 MW; 1D86A2AEF0B93EA CRC64;  
  
Query Match 87.8%; Score 36; DB 1; Length 343;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 QIIEKLVK 9  
DB 20 EVIEKLVK 28  
  
RESULT 5  
PARV\_CAEEL STANDARD; PRT; 375 AA.  
ID PARV\_CAEEL  
AC O16785;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Parvin-like protein.  
GN T21D12.4.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Woesner J.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Probably plays a role in the regulation of cell adhesion  
CC and cytoskeleton organization (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; localized to focal adhesions  
CC (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE PARVIN FAMILY.  
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AF016687; AAC48090.1; -;  
CC PIR; T28710; T28710.  
CC WormPep: T21D12.4; CEL8268.  
CC InterPro: IPR001715; Calponin-like.  
CC Pfam: PF00307; CH; 2.  
CC SMART; SM00033; CH; 2.  
CC PROSITE; PS0021; CH; 2.  
CC KW Hypothetical protein; Cell adhesion; cytoskeleton; Actin-binding;  
CC Repeat.  
CC FT DOMAIN 99 205 CH 1.  
CC FT DOMAIN 266 373 CH 2.  
SQ SEQUENCE 375 AA; 43001 MW; F9DB8B807671D0C1 CRC64;  
  
Query Match 87.8%; Score 36; DB 1; Length 375;  
Best Local Similarity 77.8%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 QIIEKLVK 9  
DB 131 QIIEKLVK 139  
  
RESULT 6  
MUS2\_BORBU STANDARD; PRT; 780 AA.  
ID MUS2\_BORBU  
AC O51125;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Mus2 protein.  
GN MUS2 OR MUTS2 OR BB0098.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Dodson R., Hickey E.K., Gwinn M.,  
RA Lathigra R., White O., Ketchum K.A., Fleischmann R.D., Richardson D.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Salzberg S., Hanson M.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Weidman J.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Wathey L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
RT burgdorferi.";  
RL Nature 390:580-586(1997).



RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward A.C., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN [3];  
RN SEQUENCE FROM N.A.  
RP TISSUE=Brain, and Urinary bladder;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan T., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smalish D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4];  
RN SEQUENCE OF 8-260 FROM N.A.  
RX MEDLINE=97368285; PubMed=9223280;  
RA Colussi P.A., Tarón C.H., Mack J.C., Orlean P.;  
RT "Human and Saccharomyces cerevisiae dolichol phosphate mannose  
RT synthases represent two classes of the enzyme, but both function in  
RT Schizosaccharomyces pombe.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:7873-7878(1997).  
RN [5];  
RN VARIANT CDG-IE GLY-92.  
RX MEDLINE=2011182; PubMed=10642597;  
RA Kim S., Westphal V., Srikrishna G., Mehta D.P., Peterson S.,  
RA Filiano J., Karnes P.S., Patterson M.C., Freeze H.H.;  
RT "Dolichol phosphate mannose synthase (DPM1) mutations define  
RT congenital disorder of glycosylation Ie (CDG-IE).";  
RL J. Clin. Invest. 105:191-198(2000).  
RN [6];  
RN VARIANT CDG-IE GLY-92.  
RX MEDLINE=2011186; PubMed=10642601;  
RA Imbach T., Schenk B., Schollen E., Burda P., Stutz A., Gruenewald S.,  
RA Ballie N.M., King M.D., Jaeken J., Matthijs G., Berger E.G., Aebi M.,  
RA Hennet T.;  
RT "Deficiency of dolichol-phosphate-mannose synthase-1 causes congenital  
RT disorder of glycosylation type Ie.";  
RL J. Clin. Invest. 105:223-239(2000).  
CC -|- FUNCTION: Transfers mannose from GDP-mannose to dolichol  
CC monophosphate to form dolichol phosphate mannose (Dol-P-Man) which  
CC is the mannose donor in pathways leading to N-glycosylation,  
CC glycosyl phosphatidylinositol membrane anchoring, and O-  
CC mannoseylation of proteins.  
CC -|- CATALYTIC ACTIVITY: GDP-mannose + dolichyl phosphate = GDP +  
CC dolichyl D-mannosyl phosphate.  
CC -|- PATHWAY: Glycosylation.  
CC -|- SUBCELLULAR LOCATION: Endoplasmic reticulum.  
CC -|- DISEASE: Defects in DPM1 are the cause of congenital disorder of  
CC glycosylation type Ie (CDG-IE) [MIM:603503]. CDG-IE is  
CC characterized by severe developmental delay, seizures, and  
CC dysmorphic features.  
CC -|- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; D86198; BAA25646.1; -;  
CC EMBL; D86202; BAA25647.1; -;  
CC EMBL; D86199; BAA25647.1; JOINED.  
CC EMBL; D86200; BAA25647.1; JOINED.  
CC EMBL; D86201; BAA25647.1; JOINED.  
CC EMBL; AL034553; CAB53749.1; -;  
CC EMBL; BC007073; AAH07073.1; -;  
CC EMBL; BC016322; AAH16322.1; -;  
CC EMBL; AF007875; AAC98797.1; -;  
CC Genew; HGNC:3005; DPM1.  
CC MIM; 603503; -;  
CC GO; GO:0004582; F:dolichyl-phosphate beta-D-mannosyltransferase. . .; TAS.  
CC GO; GO:0006486; P:protein amino acid glycosylation; TAS.  
CC InterPro; IPR001173; Glyco\_Transf\_2.  
CC Pfam; PF00535; Glycosyltransferase; Endoplasmic reticulum;  
CC Transferrase; Glycosyltransferase;  
CC Disease mutation.  
CC VARIANT 92 92 R -> G (in CDG-IE).  
CC CONFLICT 15 15 /FTID-VAR\_012341.  
CC CONFLICT 135 135 R -> W (IN REF. 4).  
CC CONFLICT 143 143 Q -> K (IN REF. 4).  
CC CONFLICT 154 154 V -> A (IN REF. 4).  
CC CONFLICT 177 177 V -> I (IN REF. 4).  
CC CONFLICT 191 191 R -> T (IN REF. 4).  
CC CONFLICT 191 191 R -> P (IN REF. 4).  
CC SEQUENCE 260 AA; 29634 MW; 9792145BFC8F0514 CRC64;  
  
CC Query Match 85.4%; Score 35; DB 1; Length 260;  
CC Best Local Similarity 66.7%; Pred. No. 28;  
CC Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
CC QY 1 QIIEKLIK 9  
CC Db 193 EVLEKLIK 201  
  
CC RESULT 9  
CC GUN\_PAPEO  
CC ID GUN\_PAPEO STANDARD; PRT; 397 AA.  
CC AC P23548;  
CC DT 01-NOV-1991 (Rel. 20, Created)  
CC DT 01-FEB-1996 (Rel. 33, Last sequence update)  
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE Endoglucoanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).  
CC OS Paenibacillus polymyxa (Bacillus polymyxa).  
CC OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
CC OX NCBI\_TaxID=1406;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=90170877; PubMed=2307659;  
CC RA Baird S.D., Johnson D.A., Sellig V.L.;  
CC RT "Molecular cloning, expression, and characterization of  
CC endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus  
CC circulans";  
CC J. Bacteriol. 172:1576-1586(1990).  
CC CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -|- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL  
CC HYDROLASES).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL: M33791; AAA22631.1; -  
 PIR: A35136; A35136.  
 HSP: P54583; LECE.  
 InterPro: IPR001547; Glyco\_hydro\_5.  
 Pfam: PF00150; cellulase; 1.  
 PROSITE: PS00659; GLYCOSYL HYDROL\_F5; 1.  
 Cellulose degradation; Hydrolase; Glycosidase.  
 ACT\_SITE 194 194 PROTON DONOR (BY SIMILARITY).  
 ACT\_SITE 317 317 NUCLEOPHILE (BY SIMILARITY).  
 SEQUENCE 397 AA; 44357 MW; B9C2E802C04FOA2A CRC64;

Query Match 85.4%; Score 35; DB 1; Length 397;  
 Best Local Similarity 77.8%; Pred. No. 41;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
 Db 127 QIMDKLEIK 135

## RESULT 10

PGK\_DROME STANDARD; PRT; 415 AA.  
 AC Q01604; Q9V0F6;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phosphoglycerate kinase (EC 2.7.2.3).  
 GN PGK OR CG3127.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon R;  
 RX MEDLINE=93101123; PubMed=1465095;  
 RA Roselli-Rehfuess L., Ye F., Lissensmore J.L., Sullivan D.T.;  
 RT "Structure and expression of the phosphoglycerate kinase (Pgk) gene  
 of *Drosophila melanogaster*.";  
 RL Mol. Gen. Genet. 235:213-220(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Ciesley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate - ADP + 3-  
 CC phospho-D-glyceroyl phosphate.  
 CC -1- PATHWAY: Second phase of glycolysis; second step.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the phosphoglycerate kinase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----

EMBL: Z14029; CAA78404.1; -  
 DR EMBL; AE003582; AAF51218.1; -  
 DR PIR; S30111; KIFFPG.  
 DR HSPP; P00560; IQPG.  
 DR FlyBase; FBgn003075; Pgk.  
 DR InterPro; IPR001576; PGK.  
 DR Pfam; PF00162; PGK; 1.  
 DR PRINTS; PR00477; PHGLYCKINASE.  
 DR PROSITE; PS00111; PGLYCERATE\_KINASE; 1.  
 DR TRANSFERASE; Kinase; Glycolysis.  
 KW CONFLICT 410 410 A -> R (IN REF. 1).  
 FT SEQUENCE 415 AA; 43861 MW; 7D1225FCC81F8B42 CRC64;  
 SQ

Query Match 85.4%; Score 35; DB 1; Length 415;  
 Best Local Similarity 66.7%; Pred. No. 42;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QIIEKLEIK 9  
 Db 262 KIVEKLEIK 270

## RESULT 11

TIG\_BUCAI STANDARD; PRT; 442 AA.  
 ID TIG\_BUCAI  
 AC P57546;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trigger factor (TF).  
 GN TIG OR B0474.  
 OS *Buchnera aphidicola* (subsp. *Acyrtosiphon pisum*) (Acyrtosiphon pisum  
 OS symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Buchnera*.  
 OX NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
*Buchnera* sp. APS.";



```

CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64002; BAA10403.1; ALT_INIT.
DR HSP; P00968; IA9X.
DR HAMAP; MF_01210; -.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPhase_L.
DR InterPro; IPR005479; CPhase_L_D2.
DR InterPro; IPR005480; CPhase_L_D3.
DR InterPro; IPR005481; CPhase_L_N.
DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPhase_L_chain; 2.
DR Pfam; PF02786; CPhase_L_D2; 2.
DR Pfam; PF02787; CPhase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRfams; TIGR01369; CPhaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 1.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 403 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 404 553 OLIGOMERIZATION DOMAIN.
FT DOMAIN 554 944 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 945 1081 ALLOSTERIC DOMAIN.
FT REPEAT 1 553
FT REPEAT 554 1081
FT NP_BIND 153 211 ATP (POTENTIAL).
FT NP_BIND 304 354 ATP (POTENTIAL).
FT METAL 286 286 MANGANESE 1 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 302 302 MANGANESE 2 (BY SIMILARITY).
FT METAL 837 837 MANGANESE 3 (BY SIMILARITY).
FT METAL 849 849 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1081 AA; 119030 MW; 0C2E0D3905B40EE6 CRC64;

Query Match 85.4%; Score 35; DB 1; Length 1081;
Best Local Similarity 66.7%; Pred. No. 99;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIEKLIK 9
Db 72 EIVEKLIK 80

RESULT 14
ACPL_CASGL STANDARD; PRT; 136 AA.
ID ACPL_CASGL
AC P93092;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Acyl carrier protein 1, chloroplast precursor (ACP 1).
GN ACPL.
OS Casuarina glauca (Swamp oak).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eurosids I; Fagales; Casuarinaceae; Casuarina.
OX NCBI_TaxID=3522;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Root nodules;
RA Lapiuze L., Gherbi H., Franche C., Duhoux E., Bogusz D.;
RT "cDNA sequence for an acyl carrier protein from actinorhizal nodules
RT of Casuarina glauca.";
RL (In) Plant Gene Register PGR98-066.
CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid
CC biosynthesis.
CC -1- PATHWAY: De novo fatty acid biosynthesis.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- PTM: 4'-phosphopantetheine is transferred from CoA to a specific
CC serine of apo-ACP by acps. This modification is essential for
CC activity because fatty acids are bound in thioester linkage to the
CC sulfhydryl of the prosthetic group (By similarity).
CC -1- SIMILARITY: Contains 1 acyl carrier domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y10994; CAA71885.1; -.
DR HSP; T09583; T09583.
DR HSP; P02901; IACP.
DR InterPro; IPR003231; Acyl_carrier.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR006162; Ppantne_attach.
DR Pfam; PF00550; pp-binding; 1.
DR ProDom; PD000887; Acyl_carrier; 1.
DR TIGRfams; TIGR00517; acyl_carrier; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
KW Fatty acid biosynthesis; Phosphopantetheine; Chloroplast;
KW Transit peptide; Multigene family.
FT TRANSIT 1 52 CHLOROPLAST (POTENTIAL).
FT CHAIN 53 136 ACYL CARRIER PROTEIN 1.
FT BINDING 91 91 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 136 AA; 14351 MW; 8DF68F53079414FB CRC64;

Query Match 82.9%; Score 34; DB 1; Length 136;
Best Local Similarity 87.5%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 IIEKLIK 9
Db 126 LIEKLIK 133

RESULT 15
LDHC_HUMAN STANDARD; PRT; 331 AA.
ID LDHC_HUMAN
AC P07864;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE L-lactate dehydrogenase C chain (EC 1.1.1.27) (LDH-C) (LDH testis
DE subunit) (LDH-X).
GN LDHC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=8726097; PubMed=2440048;
RA Millan J.L., Driscoll C.E., Goldberg E.;

```



RT "Epitopes of human testis-specific lactate dehydrogenase deduced from  
a cDNA sequence.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:5311-5315(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89193640; PubMed=2930531;  
RA Takano T., Li S.S.-L.;  
RT "Human testicular lactate dehydrogenase-C gene is interrupted by six  
introns at positions homologous to those of LDH-A (muscle) and LDH-B  
(heart) genes.";  
RT Biochem. Biophys. Res. Commun. 159:579-583(1989).  
RL [3]  
RP SEQUENCE FROM N.A.  
RX Wu K., Li S.S.-L.;  
RT "Human testicular lactate dehydrogenase-C gene: cDNA sequence and  
putative alternative splicing at the 5' noncoding region.";  
RL J. Genet. Mol. Biol. 1:72-76(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Testis;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- CATALYTIC ACTIVITY: (S)-lactate + NAD(+) = pyruvate + NADH.  
CC -!- PATHWAY: Anaerobic glycolysis; final step.  
CC -!- SUBUNIT: Homotetramer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; J02938; AAA59507.1; -  
DR EMBL; M24515; AAA59508.1; -  
DR EMBL; M24510; AAA59508.1; JOINED.  
DR EMBL; M24511; AAA59508.1; JOINED.  
DR EMBL; M24512; AAA59508.1; JOINED.  
DR EMBL; M24513; AAA59508.1; JOINED.  
DR EMBL; M24514; AAA59508.1; JOINED.  
DR EMBL; M24514; AAA59508.1; JOINED.  
DR EMBL; UL3680; AAA21348.1; -  
DR EMBL; BC019249; AAH19249.1; -  
DR PIR; A30933; DEHULC.  
DR HSP; P00339; 9LDT.  
DR Genew; HGNC:6544; LDHC.  
DR MIM; 150150; -  
DR GO; GO:0004459; F:L-lactate dehydrogenase activity; TAS.  
DR InterPro; IPR001557; L\_LDH.  
DR InterPro; IPR001236; ldh.  
DR Pfam; PF00056; ldh; 1.  
DR Pfam; PF02866; ldh\_C; 1.

DR PRINTS: PR00086; LLDHRCNASE.  
DR PROSITE: PS00064; L\_LDH; 1.  
KW Oxidoreductase; NAD; Glycolysis; Multigene family.  
FT INIT\_MET 0 0  
FT ACT\_SITE 192 192 ACCEPTS A PROTON DURING CATALYSIS.  
FT CONFLICT 76 76 V -> I (IN REF. 3).  
FT CONFLICT 118 118 S -> I (IN REF. 1).  
SQ SEQUENCE 331 AA; 36166 MW; 90A12DB15B169BC6 CRC64;  
Query Match 82.9%; Score 34; DB 1; Length 331;  
Best Local Similarity 87.5%; Pred. No. 52;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QIIEKLIE 8  
Db 6 QIIEKLIE 13  
Search completed: July 29, 2003, 18:21:23  
Job time : 12 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 17:55:15 ; Search time 95 Seconds  
(without alignments)  
24.447 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
sequence: 1 QIEKLEIK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	958	Q9QRW8	Q9qrw8 human immun
2	41	100.0	995	Q90QN7	Q90qn7 human immun
3	41	100.0	1002	Q9EAL2	Q9eal2 human immun
4	41	100.0	1003	Q90QR0	Q90qr0 human immun
5	41	100.0	1003	Q90DZ1	Q90dz1 human immun
6	41	100.0	1003	Q8JBX4	Q8jbx4 human immun
7	41	100.0	1003	Q90C17	Q90c17 human immun
8	41	100.0	1003	Q90DZ1	Q90dz1 human immun
9	41	100.0	1003	Q8JCS9	Q8jcs9 human immun
10	41	100.0	1003	Q89P59	Q89p59 human immun
11	41	100.0	1003	Q8JBS0	Q8jbs0 human immun
12	41	100.0	1003	Q8JBP2	Q8jbp2 human immun
13	41	100.0	1003	Q8JBL3	Q8jbl3 human immun
14	41	100.0	1003	Q91W52	Q91w52 human immun
15	41	100.0	1003	Q91V94	Q91v94 human immun
16	41	100.0	1003	Q89963	Q89963 human immun

17	41	100.0	1003	15	Q8AK06	Q8ak06 human immun
18	41	100.0	1272	15	Q8ADJ7	Q8adj7 human immun
19	41	100.0	1428	15	Q8ADT5	Q8adt5 human immun
20	41	100.0	1429	15	Q8AE64	Q8ae64 human immun
21	41	100.0	1430	15	Q998H8	Q998h8 human immun
22	41	100.0	1433	15	P90071	P90071 human immun
23	41	100.0	1433	15	O92651	O92651 human immun
24	41	100.0	1433	15	Q8ADV9	Q8adv9 human immun
25	41	100.0	1434	15	Q8ADT7	Q8adt7 human immun
26	38	92.7	1001	15	Q8WSF0	Q8wsf0 human immun
27	38	92.7	1002	15	Q9QST9	Q9qst9 human immun
28	37	90.2	421	17	O28350	O28350 archaeoglob
29	37	90.2	560	15	Q99B35	Q99b35 human immun
30	37	90.2	583	15	Q8QE48	Q8qe48 human immun
31	37	90.2	583	15	Q8QE44	Q8qe44 human immun
32	37	90.2	583	15	Q8QE42	Q8qe42 human immun
33	37	90.2	583	15	Q8QE46	Q8qe46 human immun
34	37	90.2	651	10	Q39720	Q39720 euglena gra
35	37	90.2	695	15	Q91V08	Q91v08 human immun
36	37	90.2	1003	15	Q9Q6X5	Q9q6x5 human immun
37	37	90.2	1003	15	Q8QDZ8	Q8qdz8 human immun
38	37	90.2	1003	15	O74453	O74453 human immun
39	37	90.2	1003	15	O90173	O90173 human immun
40	37	90.2	1003	15	Q8JBS9	Q8jbs9 human immun
41	36	87.8	366	16	Q9AAC7	Q9aac7 caulobacter
42	36	87.8	545	15	Q90S18	Q90s18 human immun
43	36	87.8	552	2	Q9FDG5	Q9fdg5 mycoplasma
44	36	87.8	564	15	Q9J4U8	Q9j4u8 human immun
45	36	87.8	999	15	Q91W43	Q91w43 human immun

ALIGNMENTS

RESULT 1

Q9QRW8 PRELIMINARY; PRT; 958 AA.  
AC Q9QRW8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Pol polyprotein (Fragment).  
GN POL.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SE8131;  
RA Laukkanen T., Salminen M.O.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL; AF107771; AAD49792.1; -  
DR HSSP; P04585; 1RT3.  
DR InterPro; IPR001995; Asparticase\_rtrv.  
DR InterPro; IPR001969; Asparticase\_site.  
DR InterPro; IPR003308; Integrase\_zn.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF02022; Integrase\_zn; 1.  
DR Pfam; PF00075; rnaaseh; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvp; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
KW Aspartyl protease; Hydrolase; Polyprotein; Protease;  
KW RNA-directed DNA polymerase; Transferase.  
FT NON\_TER 1  
SQ SEQUENCE 958 AA; 108315 MW; 1F917835072B5638 CRC64;

```

Query Match      100.0%; Score 41; DB 15; Length 958;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QIIEKLEK 9
Db 669 QIIEKLEK 677

RESULT 2
O90QN7 PRELIMINARY; PRT; 995 AA..
AC O90QN7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pol protein (Fragment).
GN POL.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21322016; PubMed=11429108;
RA Gao F., Vidal N., Li Y., Trask S.A., Chen Y., Kostrikis L.G., Ho D.D.,
RA Kim J., Oh M.-D., Choe K., Salminen M., Robertson D.L., Shaw G.M.,
RA Hahn B.H., Peeters M.;
RT "Evidence of Two Distinct Subtypes within the HIV-1 Subtype A
RT Radiation.";
RL AIDS Res. Hum. Retroviruses 17:675-688(2001).
CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC (BY SIMILARITY).
CC KNOWN AS THE RETROPEPSIN FAMILY.
EMBL: AF286240; AAK67310.1; -.
HSP; P03366; 3HYT.
InterPro: IPR001995; Aspprotease_rtrv.
InterPro: IPR001969; Aspprotease_site.
InterPro: IPR001037; Integrase_C.
InterPro: IPR003308; Integrase_Zn.
InterPro: IPR002156; RNaseH.
InterPro: IPR00477; RVTse.
Pfam: PF00552; Integrase_1.
Pfam: PF02022; Integrase_Zn; 1.
Pfam: PF00075; rve; 1.
Pfam: PF00077; rvp; 1.
Pfam: PF00078; rvt; 1.
PROSITE: PS00141; ASP_PROTEASE; 1.
PROSITE: PS0175; ASP_PROT_RETROV; 1.
Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;
Polyprotein; Protease; RNA-directed DNA polymerase; Transferase.
NON_TER 1
SQ SEQUENCE 995 AA; 113225 MW; 0537FED6756DABA CRC64;

Query Match      100.0%; Score 41; DB 15; Length 995;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QIIEKLEK 9
Db 667 QIIEKLEK 675

RESULT 3
O9EAL2 PRELIMINARY; PRT; 1002 AA.
AC O9EAL2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE Gag-pol polyprotein (Fragment).
GN POL.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21195608; PubMed=11297684;
RA Baldrich-Rubio E., Anagonou S., Stirrups K., Lafia E., Candotti D.,
RA Lee H., Allain J.P.;
RT "A complex human immunodeficiency virus type 1 A/G/J recombinant virus
RT isolated from a seronegative patient with AIDS from Benin, West
RT Africa.";
RL J. Gen. Virol. 82:1095-1106(2001).
CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC (BY SIMILARITY).
CC THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC KNOWN AS THE RETROPEPSIN FAMILY.
EMBL: AJ293865; CAC05362.1; -.
HSP; P04585; IRTH.
InterPro: IPR001995; Aspprotease_rtrv.
InterPro: IPR001969; Aspprotease_site.
InterPro: IPR001037; Integrase_C.
InterPro: IPR003308; Integrase_Zn.
InterPro: IPR002156; RNaseH.
InterPro: IPR00477; RVTse.
Pfam: PF00552; Integrase_1.
Pfam: PF02022; Integrase_Zn; 1.
Pfam: PF00075; rve; 1.
Pfam: PF00077; rvp; 1.
Pfam: PF00078; rvt; 1.
PROSITE: PS00141; ASP_PROTEASE; 1.
PROSITE: PS0175; ASP_PROT_RETROV; 1.
Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;
Polyprotein; Protease; RNA-directed DNA polymerase; Transferase.
NON_TER 1
SQ SEQUENCE 1002 AA; 113450 MW; 69687137B67359B3 CRC64;

Query Match      100.0%; Score 41; DB 15; Length 1002;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QIIEKLEK 9
Db 674 QIIEKLEK 682

RESULT 4
O90QR0 PRELIMINARY; PRT; 1003 AA.
AC O90QR0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Pol protein (Fragment).
GN POL.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=94CY017.41;
RX MEDLINE=21322016; PubMed=11429108;
RA Gao F., Vidal N., Li Y., Trask S.A., Chen Y., Kostrikis L.G., Ho D.D.,
RA Kim J., Oh M.-D., Choe K., Salminen M., Robertson D.L., Shaw G.M.,
RA Hahn B.H., Peeters M.;
RT "Evidence for two distinct subtypes within the HIV-1 subtype A
RT radiation.";
RL AIDS Res. Hum. Retroviruses 17:675-688(2001).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-94CY017.41;
RA Gao F.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC (BY SIMILARITY).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC KNOWN AS THE RETROPEPSIN FAMILY.
CC EMBL; AF286237; AAK65971.1; -.
DR InterPro; IPR001995; Aspprotease_rtrv.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase.
FT NON_TER 1
SQ SEQUENCE 1003 AA; 114111 MW; B41B6A3C95E7ECA5 CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
DB 675 QIIEKLIK 683

RESULT 5
Q90DZ1 PRELIMINARY; PRT; 1003 AA.
AC Q90DZ1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Pol protein (Fragment).
GN POL
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RA Carr J.K., Torlimiro J., Wolfe N., Eitel M., Kim B., Sanders-Buell E.,
RA Jagodzinski L., Gotte D., Burke D., Birx D., McCutchan F.;
RA "The AG Recombinant IBNG and Novel Strains of Group M HIV-1 are Common
RT in Cameroon.";
RL Virology 0:0-0(2001).
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC (BY SIMILARITY).
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC KNOWN AS THE RETROPEPSIN FAMILY.
CC EMBL; AY037284; AAL09935.1; -.
DR InterPro; IPR001995; Aspprotease_rtrv.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.

```

```

DR Pfam; PF00078; rvt; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
KW Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase.
FT NON_TER 1
SQ SEQUENCE 1003 AA; 113696 MW; 87928BD36BAAEA9 CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
DB 675 QIIEKLIK 683

RESULT 6
Q8JBX4 PRELIMINARY; PRT; 1003 AA.
AC Q8JBX4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pol protein (Fragment).
GN POL
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RA Dowling W.E., Kim B., Mason C.J., Wasunna K.Monique., Alam U.,
RA Elson L., Birx D.L., Robb M.L., McCutchan F.E., Carr J.K.;
RA "Forty-one near full length HIV-1 sequences from Kenya reveal an
RT epidemic of subtype A and A-containing recombinants.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS
CC (BY SIMILARITY).
CC EMBL; AF457064; AAN03119.1; -.
DR InterPro; IPR001995; Aspprotease_rtrv.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00552; Integrase; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; rnaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
KW Endonuclease; Hydrolase; Nucleotidyltransferase; Polyprotein;
KW RNA-directed DNA polymerase; Transferase.
FT NON_TER 1
SQ SEQUENCE 1003 AA; 113602 MW; 554C5B25EB31804D CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9
DB 675 QIIEKLIK 683

RESULT 7
Q90CI7 PRELIMINARY; PRT; 1003 AA.
AC Q90CI7;

```

DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Pol protein (Fragment).  
 GN POL.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CM53392;  
 RX MEDLINE=21342588; PubMed=11448170;  
 RA Carr J.K., Torimiro J.N., Wolfe N.D., Eitel M.N., Kim B.,  
 RA Sanders-Buell E., Jagodzinski L.L., Gotte D., Burke D.S., Birx D.L.,  
 RA McCutchan F.E.;  
 RT "The AG recombinant IBNG and novel strains of group M HIV-1 are common  
 in Cameroon";  
 RL Virology 286:168-181(2001).  
 CC -1- PFM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
 (BY SIMILARITY).  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 KNOWN AS THE RETROPEPSIN FAMILY.  
 CC EMBL: AF377957; AAK59197.1; -  
 DR InterPro; IPR001995; Asparticase\_rtrv.  
 DR InterPro; IPR001969; Asparticase\_site.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00552; Integrase\_1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; rnaaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; rvp; 1.  
 DR Pfam; PF00078; rvt; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
 KW Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;  
 KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 1003 AA; 113936 MW; 0252F04E8800CDB0 CRC64;  
 Query Match 100.0%; Score 41; DB 15; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QIIEKLEK 9  
 DB 675 QIIEKLEK 683

RESULT 8  
 ID Q90D21 PRELIMINARY; PRT; 1003 AA.  
 AC Q90D21;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE Pol protein (Fragment).  
 DE Pol protein (Fragment).  
 GN POL.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-97T202;  
 RX MEDLINE=21395692; PubMed=11504977;  
 RA Hoeleicher M., Kim B., Maboko L., Mhalu F., von Sonnenburg F.,  
 RA Birx D.L., McCutchan F.E.;  
 RA "The UNAIDS Network for HIV Isolation Characterization";  
 RT "High proportion of unrelated HIV-1 intersubtype recombinants in the  
 Mbeia region of southwest Tanzania";

RL AIDS 15:1461-1470(2001).  
 CC -1- PFM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
 (BY SIMILARITY).  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 KNOWN AS THE RETROPEPSIN FAMILY.  
 CC EMBL: AF361872; AAK94220.1; -  
 DR HSP; P03366; 3HT.  
 DR InterPro; IPR001995; Asparticase\_rtrv.  
 DR InterPro; IPR001969; Asparticase\_site.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00552; Integrase\_1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; rnaaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; rvp; 1.  
 DR Pfam; PF00078; rvt; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
 KW Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;  
 KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 1003 AA; 113661 MW; E940D5471BE95ALB CRC64;  
 Query Match 100.0%; Score 41; DB 15; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QIIEKLEK 9  
 DB 675 QIIEKLEK 683

RESULT 9  
 ID Q8JC59 PRELIMINARY; PRT; 1003 AA.  
 AC Q8JC59;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DE Pol protein (Fragment).  
 DE Pol protein (Fragment).  
 GN POL.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-00KE\_KER2009;  
 RA Dowling W.E., Kim B., Mason C.J., Wasunna K.Monique., Alam U.,  
 RA Elson L., Birx D.L., Robb M.L., McCutchan F.E., Carr J.K.;  
 RT "Forty-one near full length HIV-1 sequences from Kenya reveal an  
 epidemic of subtype A and A-containing recombinants";  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- PFM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
 (BY SIMILARITY).  
 CC EMBL: AF457053; AAN03034.1; -  
 DR InterPro; IPR001995; Asparticase\_rtrv.  
 DR InterPro; IPR001969; Asparticase\_site.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR Pfam; PF00552; Integrase\_1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; rnaaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; rvp; 1.  
 DR Pfam; PF00078; rvt; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.

DR PROSITE: PS50175; ASP\_PROT\_RETROV; 1.  
KW Endonuclease; Hydrolase; Nucleotidyltransferase; Polyprotein;  
KW RNA-directed DNA polymerase; transferase.  
FT NON\_TER 1  
SQ SEQUENCE 1003 AA; 113415 MW; 6FA94C3068E94807 CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
|||||  
Db 675 QIIEKLIK 683

RESULT 10  
O89959 PRELIMINARY; PRT; 1003 AA.  
ID O89959;  
AC O89959;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Pol protein (Fragment).  
DE Pol protein (Fragment).  
GN GAG-POL OR POL.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DJ263;  
RA Carr J.K., Salminen M.O., PubMed=9683568;  
RA Carr J.K., Salminen M.O., Albert J., Sanders-Buell E., Gotte D.,  
RA Bix D.L., McCutchan F.E.;  
RT "Full genome sequences of human immunodeficiency virus type 1 subtypes  
RT G and A/G intersubtype recombinants.";  
RL Virology 247:22-31(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DJ263;  
RA Carr J.K., Salminen M.O., Albert J., Sanders-Buell E., Gotte D.,  
RA Bix D.L., McCutchan F.E.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- PTM: SPECIFIC ENZYATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
(BY SIMILARITY).  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL; AF063223; AAC29069.1; -;  
DR HSP; P04585; IRTN.  
DR InterPro; IPR001995; Asparticase\_rtrv.  
DR InterPro; IPR001969; Asparticase\_site.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_Zn.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR Pfam; PF00552; Integrase\_1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvp; 1.  
DR Pfam; PF00078; rvt; 1.  
DR STRAIN=DJ263;  
RA Carr J.K., Salminen M.O., Albert J., Sanders-Buell E., Gotte D.,  
RA Bix D.L., McCutchan F.E.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- PTM: SPECIFIC ENZYATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
(BY SIMILARITY).  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL; AF063223; AAC29069.1; -;  
DR HSP; P04585; IRTN.  
DR InterPro; IPR001995; Asparticase\_rtrv.  
DR InterPro; IPR001969; Asparticase\_site.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_Zn.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR Pfam; PF00552; Integrase\_1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvp; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
KW Aspartyl protease; Endonuclease; Hydrolase; Nucleotidyltransferase;  
KW Polyprotein; Protease; RNA-directed DNA polymerase; transferase.  
FT NON\_TER 1  
SQ SEQUENCE 1003 AA; 113339 MW; F8D1629167DF9BEC CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9

Db 675 QIIEKLIK 683  
|||||

RESULT 11  
Q8JBS0 PRELIMINARY; PRT; 1003 AA.  
ID Q8JBS0;  
AC Q8JBS0;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Pol protein (Fragment).  
GN POL.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=00KE\_KNH1211;  
RA Dowling W.E., Kim B., Mason C.J., Wasunna K.Monique., Alam U.,  
RA Elson L., Bix D.L., Robb M.L., McCutchan F.E., Carr J.K.;  
RT "Forty-one near full length HIV-1 sequences from Kenya reveal an  
RT epidemic of subtype A and A-containing recombinants.";  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- PTM: SPECIFIC ENZYATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
(BY SIMILARITY).  
DR EMBL; AF457070; AAN03173.1; -;  
DR InterPro; IPR001995; Asparticase\_rtrv.  
DR InterPro; IPR001969; Asparticase\_site.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_Zn.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR Pfam; PF00552; Integrase\_1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvp; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
KW Endonuclease; Hydrolase; Nucleotidyltransferase; Polyprotein;  
KW RNA-directed DNA polymerase; transferase.  
FT NON\_TER 1  
SQ SEQUENCE 1003 AA; 113662 MW; D65AA71EASF16124 CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLIK 9  
|||||

Db 675 QIIEKLIK 683

RESULT 12  
Q8JBP2 PRELIMINARY; PRT; 1003 AA.  
ID Q8JBP2;  
AC Q8JBP2;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Pol protein (Fragment).  
GN POL.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=99KE\_KSM4021;  
RA Dowling W.E., Kim B., Mason C.J., Wasunna K.Monique., Alam U.,  
RA Elson L., Bix D.L., Robb M.L., McCutchan F.E., Carr J.K.;

RT "Forty-one near full length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A'-containing recombinants."  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
 CC (BY SIMILARITY)."

DR EMBL: AF457075; AA03201.1; -.  
 DR InterPro: IPR001995; Aspprotease\_rtrv.  
 DR InterPro: IPR001969; Aspprotease\_site.  
 DR InterPro: IPR001037; Integrase\_C.  
 DR InterPro: IPR003308; Integrase\_Zn.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR000477; RVTse.  
 DR Pfam: PF00552; Integrase\_1.  
 DR Pfam: PF02022; Integrase\_Zn; 1.  
 DR Pfam: PF00075; rnaseH; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR Pfam: PF00077; rvp; 1.  
 DR Pfam: PF00078; rvt; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE: PS0175; ASP\_PROT\_RETROV; 1.  
 KW Endonuclease; Hydrolase; Nucleotidyltransferase; Polyprotein;  
 FT RNA-directed DNA polymerase; Transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 1003 AA; 113509 MW; B452CCE0B8D2F25B CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9  
 | | | | | | | | | |  
 Db 675 QIIEKLEIK 683

RESULT 13

ID Q8JBL3 PRELIMINARY; PRT; 1003 AA.  
 AC Q8JBL3;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Pol protein (Fragment).  
 GN POL.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrov. viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=00KE\_KSM4030;  
 RA Dowling W.E., Kim B., Mason C.J., Wasunna K.Monique., Alam U.,  
 RA Elson L., Birx D.L., Robb M.L., McCutchan F.E., Carr J.K.;  
 RT "Forty-one near full length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A'-containing recombinants."  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
 CC (BY SIMILARITY).  
 DR EMBL: AF457079; AA03230.1; -.  
 DR InterPro: IPR001995; Aspprotease\_rtrv.  
 DR InterPro: IPR001969; Aspprotease\_site.  
 DR InterPro: IPR001037; Integrase\_C.  
 DR InterPro: IPR003308; Integrase\_Zn.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR001584; Rve.  
 DR Pfam: PF00552; Integrase\_1.  
 DR Pfam: PF02022; Integrase\_Zn; 1.  
 DR Pfam: PF00075; rnaseH; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR Pfam: PF00077; rvp; 1.  
 DR Pfam: PF00078; rvt; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE: PS0175; ASP\_PROT\_RETROV; 1.

KW Endonuclease; Hydrolase; Nucleotidyltransferase; Polyprotein;  
 KW RNA-directed DNA polymerase; Transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 1003 AA; 113468 MW; F017CAAF51AF0B5E CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIIEKLEIK 9  
 | | | | | | | | | |  
 Db 675 QIIEKLEIK 683

RESULT 14

ID Q9IW52 PRELIMINARY; PRT; 1003 AA.  
 AC Q9IW52;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE POL protein.  
 GN POL.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrov. viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=98SE-MP1211;  
 RA Montavon C.;  
 RA Montavon C.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=98SE-MP1211;  
 RX MEDLINE=20236909; PubMed=10777151;  
 RA Toure-Kane C., Montavon C., Faye M., Gueye P., Sow P., Ndoeye I.,  
 RA Gaye-Diallo A., Delaporte E., Peeters M., Mboup S.;  
 RT "Identification of all HIV type 1 group M subtypes in Senegal, a country with low and stable seroprevalence."  
 RT AIDS Res. Hum. Retroviruses 16:603-609(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=98SE-MP1211;  
 RA Montavon C., Toure-Kane C., Liegeois F., Mpoudi E., Bourgeois A.,  
 RA Vergne L., Perret J.L., Boumah A., Saman E., Mboup S., Delaporte E.,  
 RA Peeters M.;  
 RT "Most env and gag subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG.";  
 RT J. Acquir. Immune Defic. Syndr. 15:363-374(2000).  
 RL -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
 CC (BY SIMILARITY).  
 DR EMBL: AJ251056; CAB86366.1; -.  
 DR HSP: P03366; IHV.  
 DR InterPro: IPR001995; Aspprotease\_rtrv.  
 DR InterPro: IPR001969; Aspprotease\_site.  
 DR InterPro: IPR001037; Integrase\_C.  
 DR InterPro: IPR003308; Integrase\_Zn.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR000477; RVTse.  
 DR Pfam: PF00552; Integrase\_1.  
 DR Pfam: PF02022; Integrase\_Zn; 1.  
 DR Pfam: PF00075; rnaseH; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR Pfam: PF00077; rvp; 1.  
 DR Pfam: PF00078; rvt; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE: PS0175; ASP\_PROT\_RETROV; 1.  
 KW Endonuclease; Hydrolase; Nucleotidyltransferase; Polyprotein;  
 KW RNA-directed DNA polymerase; Transferase.  
 SQ SEQUENCE 1003 AA; 113446 MW; AB5F5FA43D9EC5FC CRC64;



Query Match 100.0%; Score 41; DB 15; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIEKLIK 9  
| | | | | | | | | |  
DB 675 QIEKLIK 683

RESULT 15  
Q9IV94 PRELIMINARY; PRT; 1003 AA.  
AC Q9IV94;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Pol polyprotein (Fragment).  
GN POL.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VI1197;  
RX MEDLINE=20418941; PubMed=10954895;  
RA Janssens W., Salminen M.O., Laukkanen T., Heyndrickx L.,  
RT "Near full-length genome analysis of HIV type 1 CRF02\_AG subtype C and  
RT CRF02\_AG subtype G recombinants";  
RL AIDS Res. Hum. Retroviruses 16:1183-1189(2000).  
CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
DR EMBL; AJ276596; CAB82227.1; -.  
DR HSP; P04585; IRT3.  
DR InterPro; IPR001995; Asparticase\_rtrv.  
DR InterPro; IPR001969; Asparticase\_site.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_Zn.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00552; Integrase; 1.  
DR Pfam; PF02022; Integrase\_Zn; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvp; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
KW Aspartyl protease, Endonuclease, Hydrolase, Nucleotidyltransferase;  
KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase.  
FT NON\_TER 1  
SQ SEQUENCE 1003 AA; 113117 MW; 4F80E83AFD090DDC CRC64;

Query Match 100.0%; Score 41; DB 15; Length 1003;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QIEKLIK 9  
| | | | | | | | | |  
DB 675 QIEKLIK 683

Search completed: July 29, 2003, 18:23:07  
Job time : 98 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 6, 2003, 07:06:41 ; Search time 2084 Seconds  
(without alignments)  
176.673 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIEKLEIK 9

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xl  
-Q/cgn2\_1/USPTO.spool/US0994617/runat\_29072003\_160422\_7634/app\_query.fasta\_1.199  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0994617 @CNG.1.1.5265 @runat\_29072003\_160422\_7634 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htg\_mus:\*  
34: em.htg\_pln:\*  
35: em.htg\_rod:\*  
36: em.htg\_mam:\*  
37: em.htg\_vrt:\*  
38: em.sy:\*  
39: em.htgo\_hum:\*  
40: em.htgo\_mus:\*  
41: em.htgo\_other:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	41	100.0	3012	6	AR264808	Sequence
2	41	100.0	3569	12	AY227361S1	AY227361 Synthetic
3	41	100.0	5100	14	AF286240	AF286240 HIV-1 iso
4	41	100.0	5191	14	AY037284S1	AY037284 HIV-1 ffo
5	41	100.0	5323	14	AF286241	AF286241 HIV-1 iso
6	41	100.0	8277	14	AY151002	AY151002 HIV-1 iso
7	41	100.0	8389	14	AY093607	AY093607 HIV-1 iso
8	41	100.0	8675	14	AF457074	AF457074 HIV-1 iso
9	41	100.0	8757	14	HIM293865	AF293865 Human imm
10	41	100.0	8771	14	AF457079	AF457079 HIV-1 iso
11	41	100.0	8788	14	AF484517	AF484517 HIV-1 iso
12	41	100.0	8794	14	AF484496	AF484496 HIV-1 iso
13	41	100.0	8796	14	AF484479	AF484479 HIV-1 iso
14	41	100.0	8803	14	AY093603	AY093603 HIV-1 iso
15	41	100.0	8806	14	AF457060	AF457060 HIV-1 iso
16	41	100.0	8811	14	AF457070	AF457070 HIV-1 iso
17	41	100.0	8822	14	AF457075	AF457075 HIV-1 iso
18	41	100.0	8829	14	AF361872	AF361872 HIV-1 iso
19	41	100.0	8833	14	AF457053	AF457053 HIV-1 iso
20	41	100.0	8835	14	AF484493	AF484493 HIV-1 iso
21	41	100.0	8846	14	AF457064	AF457064 HIV-1 iso
22	41	100.0	8851	14	HIM251056	AF251056 Human imm
23	41	100.0	8853	14	AF484507	AF484507 HIV-1 iso
24	41	100.0	8896	14	AF377957	AF377957 HIV-1 iso
25	41	100.0	8914	14	HIM276596	AJ276596 HIV-1 pro
26	41	100.0	8961	14	AF063224	AF063224 HIV-1 iso
27	41	100.0	8972	14	AF286238	AF286238 HIV-1 iso
28	41	100.0	8999	14	HIVU51190	U51190 HIV-1 clone
29	41	100.0	9002	14	AF063223	AF063223 HIV-1 iso
30	41	100.0	9016	14	AF107771	AF107771 HIV-1 iso
31	41	100.0	9039	14	AY093605	AY093605 HIV-1 iso
32	41	100.0	9060	6	AR264728	AR264728 Sequence
33	41	100.0	9060	14	AF286237	AF286237 HIV-1 iso
34	41	100.0	9160	14	AF004885	AF004885 HIV-1 iso
35	41	100.0	9178	14	HIVU455A	M62320 Human immu
36	41	100.0	9397	14	AF197341	AF197341 HIV-1 iso
37	41	100.0	9628	14	AF197340	AF197340 HIV-1 iso
38	41	100.0	9708	14	AB052867	AB052867 Human imm
39	41	100.0	9748	14	AB049811	AB049811 Human imm
40	41	100.0	9843	14	HIVU51188	U51188 HIV-1 clone
41	39	95.1	4851	6	E33012	E33012 Primer for
42	39	95.1	39377	3	U37430	U37430 Caenorhabdi
43	39	95.1	135912	9	AC092208	AC092208 Homo sapi
44	38	92.7	8813	14	AF071474	AF071474 HIV-1 iso
45	38	92.7	8929	14	AF076474	AF076474 HIV-1 iso

# ALIGNMENTS

RESULT 1

```

AR264808
LOCUS AR264808 3012 bp DNA 1linear PAT 10-APR-2003
DEFINITION Sequence 92 from patent US 6492110.
ACCESSION AR264808
VERSION AR264808.1 GI:29693174
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE 1 (bases 1 to 3012)
AUTHORS Hahn,B.H., Shaw,G.M. and Gao,F.
TITLE Reference clones and sequences for non-subtype B isolates of human immunodeficiency virus type 1
JOURNAL Patent: US 6492110-A 92 10-DEC-2002;
FEATURES
    Location/Qualifiers
    1..3012
        /organism="unknown"
BASE COUNT 1186 a 496 c 676 g 654 t
ORIGIN
Alignment Scores:
Pred. No.: 23.5 Length: 3012
Score: 41.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-994-617-4 (1-9) x AR264808 (1-3012)

QY 1 GlnlleleGluLysLeulleGluLys 9
|||||
Db 2023 CAATAATAGAAAATAATAGAAAAG 2049

RESULT 2
AY227361S1
LOCUS AY227361S1 3569 bp RNA 1linear SYN 02-APR-2003
DEFINITION Synthetic construct HIV-1 derived gag protein (gag) gene, complete cds; and pol protein (pol) gene, partial cds.
ACCESSION AY227361
VERSION AY227361.1 GI:29470216
KEYWORDS
SEGMENT 1 of 2
SOURCE Synthetic construct
ORGANISM Synthetic construct
AUTHORS Ellenberg,D.L., Li,B., Smith,J.M., Yi,H., Folks,T.M., Robinson,H.L. and Butera,S.T.
TITLE Optimization of a multigene HIV-1 recombinant subtype CRF02_AG DNA vaccine for expression of multiple immunogenic forms
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3569)
AUTHORS Ellenberg,D.L., Li,B., Lupo,L.D., Smith,J.M., Nkengasong,J., Radio-Morokro,M.S., Robinson,H.L., Ackers,M.-L., Greenberg,A.E., Folks,T.M. and Butera,S.T.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) NCID/DASTLR/HARB, Centers for Disease Control and Prevention, 1600 Clifton Rd., MS G-19, Atlanta, GA 30333, USA
FEATURES
    Location/Qualifiers
    1..3569
        /organism="synthetic construct"
        /mol_type="genomic RNA"
        /db_xref="taxon:32630"
        /note="derived from Human immunodeficiency virus 1 recombinant subtype CRF02_AG isolate IC0928 from Cote d'Ivoire"
    98..1585
        /gene="gag"
    98..1585
        /gene="gag"
        /codon_start=1

gene
CDS

```

```

/transl_table=11
/product="gag protein"
/protein_id="AA074589.1"
/db_xref="GI:29470218"
/transl_table="NGARASVLTGGKLDSEKIRLRPGGKKRYRLKHLVWASRELERF
ALNPGLETAEGCOQLEQIPALRTGSEEFKSLHNTVATLWCHORIDIKDQEAID
KVEEKQNSKQAKQAATAATAGSSQNYPIVQNAQGMVHQSMSPTLNAWVKVIE
EKAFSPVIFPMFSALSEGATPOLNMLNI VGGHQAOMQKDTINDEAAEWDVHPV
HAGPIPPGMREPRGSDIAGTTSTLOEIGWMTSNPIPVGEIYKWI VGLNKLIVRM
YSPSTILDIRQPKPRDYVDFEFTLRAEQATQEVKNWMTETLLVQNAPOCKSIL
RALPGATLEEMMTSCQGGPGHKARVLAEMSQVOQTVMQGNFRGQRIKCFN
CGREGLHARNCAPRKRCWKCGKQGMKQDCTERQANFLGKTPWSHKGRPGNQSR
PEPTAPAESFGVGEETPSPKQPRDKGLYPLPLTSLKSLFGNDQ"
<1387..>3569
/gene="pol"
<1387..>3569
/gene="pol"
/codon_start=1
/transl_table=11
/product="pol protein"
/protein_id="AA074590.1"
/db_xref="GI:29470219"
/transl_table="FERNLAFPQGEARKFSSQRTTNSPASRELHSGGRDITLLSEAG
AEGGTSSLNFPQITLWQRPVTVVTRIGGQPIEALDGTGADDTVLEEISLPGKWKPM
IGGIGFIRVQYDQISIEICGKRAIGTVLGPVTVNIIGRNMLTQIGCTLNPISPI
ETVPVKLSGMDGPKVQWPLTEEKIKALKEICAEMEKEGKISKIGPENYNTPIFAI
KKDSTKWRKLVDRELNRKRTQDFWVQLGIPHPAGLLKKKSVTVLDVGDAYEVPLO
EDFRKYATPIPSLNNETGIRYQVNVLPQGWKSPAI FOASMTKILEPFRANKPEIV
IYQMDLLYVGSDELEIGHRAKIEELREHLLKWGTFPPDKKHQKEPFLWGLMHPD
KWTVPQLPQKDSWTVNDIQKLVGLNWSAQIYAGIKVKQCLRLURGAKALTDVVTL
TEARELAENREILKEPVHGVIYDPTKDLVAEIQKQGDQWYIYQEPFKMLTKG
YAKRSAAHTNDVKQLTEVQKIAIESIVIMGKTPKFLPIQRETWEAMWMEYQATWI
PEWFEVNTPLVQLTQLEKDPIMGAETFYVDGAANRETKLGKAGVYVDRGKQVSL
IETNOKTELHATHLADSGSEVNIIVTDSQVAGIIOAQPDSESESLVNIIEKLE
KDKVYLSWVPAHKGIGNEQVDLVSSGIRKVLFLDGIKQADEH"
BASE COUNT 1361 a 617 c 848 g 743 t
ORIGIN
Alignment Scores:
Pred. No.: 28.4 Length: 3569
Score: 41.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-09-994-617-4 (1-9) x AY227361S1 (1-3569)

QY 1 GlnlleleGluLysLeulleGluLys 9
|||||
Db 3409 CAAATAATAGAAAATAATAGAAAAG 3435

RESULT 3
AF286240
LOCUS AF286240 5100 bp DNA 1linear VRL 27-JUN-2001
DEFINITION HIV-1 isolate 97CDKFE4 from Republic of the Congo gag protein (gag) and pol protein (pol) genes, partial cds.
ACCESSION AF286240
VERSION AF286240.1 GI:14571890
KEYWORDS
SOURCE Human immunodeficiency virus 1 (HIV-1)
ORGANISM Human immunodeficiency virus 1
REFERENCE 1 (bases 1 to 5100)
AUTHORS Gao,F., Vidal,N., Li,Y., Trask,S.A., Chen,Y., Kostrikis,L.G., Ho,D.D., Kim,J., Oh,M.-D., Choe,K., Salminen,M., Robertson,D.L., Shaw,G.M., Hahn,B.H. and Peeters,M.
TITLE Evidence of two distinct subtypes within the HIV-1 subtype A radiation
JOURNAL AIDS Res. Hum. Retroviruses 17 (8), 675-688 (2001)
MEDLINE 21322016
PUBMED 11429108

```























```
/translation="MEQAPDQGPQREPYNAAWELLEELKSEAVRHFPVRLHGLGQ
HIVTYGDTWTGVEAIRILQLLFIHYRICQHSRIGITPQRTNRGSSRS"
5020..7631
/gene="tat"
Join(5020..5234,7586..7631)
/gene="tat"
/codon_start=1
/product="tat protein"
/protein_id="AAN73604.1"
/db_xref="GI:25166814"
/translation="MDPVDNLEPNHHPGSQPESPCNKCYCKKCYHCQLCFITKGLG
ISYGRKRRQRRTQSDAQHQVPIPNQSPSSQPRGDTGPKE"
5159..7869
/gene="rev"
Join(5159..5234,7586..7869)
/gene="rev"
/note="protein extends 3 amino acids beyond HXB2 Rev"
/codon_start=1
/product="rev protein"
/protein_id="AAN73605.1"
/db_xref="GI:25166815"
/translation="MAGRSGDSELLKVIPLIKELVQTNPSPSEGTQARRNRNR
WRKQRQHSIGERLIISYLGRRPEPVLPPLERLSLNCNEDCGTSGTGUGVSPQI
LVESPAVLDSGTKECY"
5251..5505
/gene="vpu"
5251..5505
/gene="vpu"
/codon_start=1
/product="vpu protein"
/protein_id="AAN73606.1"
/db_xref="GI:25166816"
/translation="MQSIVILAIIVAVFLIAIVMTIVIECKRKLQKQKRWDLIK
BISRAEDSGNESGDEGTAELSRLVEMGNLGYNDL"
5414..5554
/gene="env"
5414..5554
/gene="env"
/note="gp160; cleaved into signal, gp120, and gp41;
protein truncated by premature stop codon relative to HXB2
gp160"
/codon_start=1
/product="truncated envelope glycoprotein"
/protein_id="AAN73607.1"
/db_xref="GI:25166817"
/translation="MRVMARARQBNYPDLRWGTIILGMINICSTAEIDLWTVYGV
PV"
8004..8618
/gene="nef"
8004..8618
/gene="nef"
/codon_start=1
/product="nef protein"
/protein_id="AAN73608.1"
/db_xref="GI:25166818"
/translation="MGGKWSKSSIVGWPEVRIRRTPAAGKGVGAVSQDLDKCGAIT
SSNVNHPSCVLEAQDEDEGVFPVPOVPLRPMYKAAFDSLHFKEGGLGIYSR
KROELDLWVHTGGFFDQWQNYTGGTFRPLFGWCFLKVPVNDVVERATEGNT
SILHPISGHMDDEEGTWEFDSRLAKHRAQEMHPEFYKDC"
8287..>8794
/note="complete U3 and partial R-repeat region"
BASE COUNT 3175 a 1564 c 2076 g 1979 t
ORIGIN

Alignment Scores:
Pred. No.: 76.2 Length: 8794
Score: 41.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-994-617-4 (1-9) x AF484496 (1-8794)
```

```
QY 1 GlnlleleGluLysLeulleGluLys 9
|||||
Db 3297 CAATAATAGAGAGAGCTAATAAGAAA 3323

RESULT 13
AF484479 8796 bp DNA linear VRL 22-NOV-2002
HIV-1 isolate 99UGC06443 from Uganda, partial genome.
AF484479
VERSION AF484479.1 GI:25166648
KEYWORDS
SOURCE
ORGANISM Human immunodeficiency virus 1 (HIV-1)
Viruses; Retroïd viruses; Retroviridae; Lentivirus; Primate
lentivirus group.
REFERENCE 1 (bases 1 to 8796)
AUTHORS Harris,M.B., Servadda,D., Sewankambo,N., Wabwire,F., Kim,B.,
Kigozi,G., Kiwanuka,N., Phillips,J.B., Meehen,M., Lutalo,T.,
Lane,J.R., Merling,R., Gray,R., Wawer,M., Birx,D.L., Robb,M.L. and
McCutchan,F.E.
TITLE Among 46 Near Full Length HIV Type 1 Genome Sequences from Rakai
JOURNAL District, Uganda, Subtype D and AD Recombinants Predominate
REFERENCE 2 (bases 1 to 8796)
AUTHORS Gray,R.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2002) Johns Hopkins University, Baltimore, MD,
USA
REFERENCE 3 (bases 1 to 8796)
AUTHORS Harris,M.B., Birx,D.L. and Robb,M.L.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2002) Walter Reed Army Institute of Research,
Washington, DC, USA
REFERENCE 4 (bases 1 to 8796)
AUTHORS Kim,B., Phillips,J.B., Lane,J.R., Merling,R. and McCutchan,F.E.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2002) The Henry M. Jackson Foundation, Rockville,
MD, USA
REFERENCE 5 (bases 1 to 8796)
AUTHORS Lutalo,T.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2002) Uganda Virus Research Institute, Entebbe,
Uganda
REFERENCE 6 (bases 1 to 8796)
AUTHORS Meehen,M. and Wawer,M.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2002) Columbia University, New York, NY, USA
REFERENCE 7 (bases 1 to 8796)
AUTHORS Servadda,S., Sewankambo,N., Wabwire,F., Kigozi,G. and Kiwanuka,N.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2002) Makerere University, Kampala, Uganda
FEATURES
source
location/Qualifiers
1..8796
/organism="Human immunodeficiency virus 1"
/proviral
/mol_type="genomic DNA"
/isolate="99UGC06443"
/db_xref="taxon:11676"
/country="Uganda: Rakai"
/note="subtype: AD"
<1..4289
/gene="gag-pol"
Join(<1..1284,1284..4289)
/gene="gag-pol"
/note="translated by -1 ribosomal slip on ttttt slippary
sequence followed by inverted repeat stem-loop structure
which potentiates slippage; ribosomal slippage"
/codon_start=1
/product="gag-pol fusion polypeptide"
/protein_id="AAN73456.1"
/db_xref="GI:25166649"
/translation="ARASVLSGGKLDWEKIRLRPGKKYRLKHLVWASRELERFAL
```





```
|||||
Db      3300 CAAATAATAGAGAAATTAATAGAAAA 3326

RESULT 14
AY093603 LOCUS      8803 bp      DNA      linear      VRL 07-APR-2003
DEFINITION HIV-1 isolate 95SN1795 from Senegal, complete genome.
ACCESSION  AY093603
VERSION     AY093603.1 GI:29409294
KEYWORDS
SOURCE      Human immunodeficiency virus 1 (HIV-1)
ORGANISM    Human immunodeficiency virus 1
Virus; Retrovirdae; Retroviridae; Lentivirus; Primate
lentivirus group.
1 (bases 1 to 8803)
McCutchan,F.E., Sankale,J.-L., M'Boup,S., Kim,B., Tovanabutra,S.,
Hamel,D.J., Brodine,S., Kanki,P.J. and Birx,D.L.
HIV-1 Circulating Recombinant Form CRF09_cpx from West Africa
combines subtypes A, F, G, and may share ancestors with CRF02_AG
and Z321
Unpublished
2 (bases 1 to 8803)
McCutchan,F.E., Sankale,J.-L., M'Boup,S., Kim,B., Tovanabutra,S.,
Hamel,D.J., Brodine,S., Kanki,P.J. and Birx,D.L.
Direct Submission
JOURNAL
SUBMITTED (28-MAR-2002) Global Molecular Epidemiology, Henry M.
Jackson Foundation, 1 Taft Court, Rockville, MD 20850, USA
Genome sequence lacks part of non-coding region.
Location/Qualifiers
1..8803
/organism="Human immunodeficiency virus 1"
/proviral
/mol_type="genomic DNA"
/isolate="95SN1795"
/isolation_source="primary PBMC of a 36 yr. old female CSW
who seroconverted to HIV-1 in 1995; sample was obtained
within one year of seroconversion"
/db_xref="taxon:11676"
/country="Senegal"
1..1494
/gene="gag"
1..1494
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="AA067375.1"
/db_xref="GI:29409295"
/translation="MGARASVLSGKLDKWEKIRLPGKKQYKMKHLVWASRELERF
AVNPSLLEAECQOIMQALPALTGSEELSLYNTVATLYCVHORIDVDRDKALD
KTEIQKSKQETQAAANTGSNSKVSQNYPIVNAQGMVHQAI SPTLNAWVKVIE
EKAFSPVIMFTALSEGATPDLMNMLNIVGGHQAAMQMLKDTINEEAVEDWTRHPV
HAGNPPGCMREPRGSDIAGTISNPQEQIAWNTGPPVPGVEIKRWIVGLNKLIVRM
YSPVILIRGPKPEPPFDYVDRFKTLRAEQASOEVRNMTDTLLQVNAAPDCKTIL
KALGPATLEEMTACQGVGSHKARVLAAMSQAQNTNIMORSNFKGQRIKCFN
CGEGHLARNCRKPCGKCGHQMKDCTERQANFLGKIWP SHKRGPNFPQSR
PEPAPPAEALGMBREITSSLKQEGGQSPPLVSLKSLFGNDPLSQ"
<1290..4298
/gene="pol"
<1290..4298
/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="AA067375.1"
/db_xref="GI:29409296"
/translation="FFREDLAPQGEARELSSEQTGAISPTSRELGDEGRNLLSEAG
EGGTVSFSFPQITLWQRPVTVKIGQLKEALDGTADDTLVEINLPGKWKPKMI
GGIGFIVKRVQDQPIEICGORAIGTVGTPVNIIGRNMLTOIGCTLNPISPIE
TPVVKLPGMDGPKTKOWPLTEETKALKEICTEMEKEKISRVCENPYNTPVFAIK
KDSKWKRLVDFELNKLKRTQDFWEVQLGPHPAGLKKKSVTVLDVGDATFSVPLDK
DFRKTTATFSTNETFGIRIQYNLVPGWKGSPAIQSSMRILEFRIRKNPEIVI
YOYMDLDTSDTEIGOHRATIEELRHLKMGFTTPDKKHQKEPPFLWMGYELHPDK
WTQVPIQIPKESWTVDNIQKLVKNWASQIYAGIKVRQLCKLRGAKALTDIVPLT
BEAELELAENREILKEPVHGVYDPKDLVAIEIQKQGDQWTYQIYQEPFNKLTGY
```

```
AKKSAHTNDVRQLTEVVOKIATESIVIGTKPKFKLPQOKETWETWADYQWASWIP
EWEFVNTPLVLYQLBKDPIIAGAEETVVDCAANRETKLKGAGVTVNRGRKVSLSLT
ETNOKTELHAIYALQDSGSEVNIIVTSQYALGIIQAQDRSDSEIYNQIEKLEIK
EKVYLSWVPAHKGIGGQVSKVIRKVLFDGIDKRAQEEHRYHNSWRAMASDF
NLPIVAKIEIVASCDKQKLGAEVQDQSPGIWQIDCTHLEKGIILVAVHVASGII
EAEVIPAETGETAYFILKLAGRWPKVHTDNGSNFTSAAVKAAACWAGITQEEGIP
YNPQSGVSNMKNELKIIIGQVRQAEHLKTAQVMVFIHNFKRKGGIGEYSAGERI
VDIIASDIQTELKQITKIQNFRVYRDSRDPWKGPAPKLLWKEGAVVIQDKSDIK
VVPREKAKIIRHYGKQMGAGDDCVAGQDED"
4243..4821
/gene="vif"
4243..4821
/gene="vif"
/codon_start=1
/product="vif protein"
/protein_id="AA067377.1"
/db_xref="GI:29409297"
/translation="MENRWQVMVIMVQVDRMINTWKSIVKYHMHISKKAKKWFYRHYY
DSNHPKISSEVHIPLGDAELVVTYGLHTGEREHLGQGVSIERLKKRYRTQVDPDL
ADOLTHIYFDCFTESAIRKAILGHRVSPRCYQAGHPKVGSLQYLAUTALIKPKTK
PPLPSVKKLIVEDRWNKPQTRGRRENQTMNGH"
4761..5054
/gene="vpr"
4761..5054
/gene="vpr"
/codon_start=1
/product="vpr protein"
/protein_id="AA067378.1"
/db_xref="GI:29409298"
/translation="MEQAPEDQGPQEPNNEWTLEELKNEAVKHFPRLWHELGO
HIYATYGDTEGVCQAIIRILQQLLEVFHFRIGCHHSRIGIIRRRVRDAGRS"
5035..7668
/gene="tat"
join(5035..5246,7566..7668)
/gene="tat"
/codon_start=1
/product="tat protein"
/protein_id="AA067379.1"
/db_xref="GI:29409299"
/translation="MEPVDNLEPNHFGSRPATACNSCYKCKCYHCQKCFNLKGLG
ISYGRKKRRQRRAAGSGSKSNODLIPNCRSRPLSQT RGDPTGPKESKKEVESKTKTDP
FD"
5174..7852
/gene="rev"
join(5174..5246,7566..7852)
/gene="rev"
/codon_start=1
/product="rev protein"
/protein_id="AA067380.1"
/db_xref="GI:29409300"
/translation="MAGRSGDSEOLLAAVKAIKILYQIVVDPDPYKPGTQARNR
RRWRARQTHSISERILSTCLGRLEPVLQPLPRLRLHLDCESDCGNSGTQGVAD
PQISGESCTVLESQTK"
5276..5521
/gene="vpu"
5276..5521
/gene="vpu"
/codon_start=1
/product="vpu protein"
/protein_id="AA067381.1"
/db_xref="GI:29409301"
/translation="MQSLYIASIVGLIIVAFIAAIIWTIVFIEVRKIRQKRIDLRI
RIRDRAESGNESEGDAAELAEVEMGHDLNWNADL"
5439..7994
/gene="env"
5439..7994
/gene="env"
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA067382.1"
/db_xref="GI:29409302"
/translation="MRVKGMRNQNLRWGLMIRGMLLICNAADNLWTVTYGVPPVW
KDAETTLFCASDAKAYSTEKHNVWATHACVPTDPNPNQELGNTVENNMKNMVEQ
```

```
MHEDIISLMDQSLRCPVKLTPLCVTLDCNDVSRNNTNNSSIFSNTEDGRGEMANCSF
NMTEURDKQRVHSLFYRLDIVKINETENNRSHYRLINCNSTSVITQACPKVTFEPI
PIHYCAPAGFAILKCRKFKNGTGTNNVSVQCTHGICKPWSTQLLNGSLAEEDII
IRENTISNAKVIIVOLDKPKVKNCTRPNNTRKGVHIGPGQAFYATDDIVGDIRAF
CEVNTQWNETLAKVEKLEKHEFNKTIISQPSGGDLEITTHSFNGCGGFCYNTN
LFWNNTKDIMPNGTANYFOCKIKQIWRMVRQGMAYAPPPIGPIECVSNITGLVL
TROGGTSETFPTFGGNNKNRSELYKYVYKIEPLGVAPTRAKRRYVEREKRAV
GMALFGLFGAAGSTMGASMTLTVAQRLSGIVQOQSNLLKATEAQHLLKLTWV
GIKQALAVLARYLKDOOLLIGWCSKGLICTVNPWNSMKNTOOIBWMTWL
EMDREISNTDIYIKLIESQNOEKNEODLLALDSWASLWNEFDITRLWYIKITM
IVGGLGIRIIFAVLSVWVRGYSPLSPQTLIPNRPDPREIEEGGEGDCKERS
IRVLSFGLAWDDLNLCLFSTRHRLDFILIAARTVEILGHRVQWLKYLENLAQYW
SOEKNSAISLLNTAIVVAEGTDRIETVQLORAGRAILHTPRIROGLERALI"
/sequence .8616
/organism="nef"
/length .8616
/organism="nef"
/codon_start=1
/product="nef protein"
/protein_id="AA067383.1"
/db_xref="GI:29409303"
/translation="MGKWSKGIWGPETRIIRKSPAAEGVAVSODLAKHGAI
SSNTASTNPDCAMLEQEDVDGPPVPOVPLRPMTEKGAFLDSFLKEKGGLEGLY
SRKRRLDLWYHTQGYPDQWNTYEGVREPLTEGRCFKLVPYDPKVEVEATGGE
NNCLLHPMCOHGIEDSEGEILMKFSTILAKHLAKHKEHPEYTKDC"
BASE COUNT 3230 a 1516 c 2108 g 1949 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 76.3 Length: 8803
Score: 41.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
```

US-09-994-617-4 (1-9) x AY093603 (1-8803)

```
Qy 1 GlnlelleGlulysLeulleGlulys 9
|||||
Db 3309 CAAATAATAGAGAAGCTAATAGAAAA 3335
```

## RESULT 15

```
AF457060
LOCUS AF457060 8806 bp DNA linear VRL 11-OCT-2002
DEFINITION HIV-1 isolate 00KE_KISII5009 from Kenya, partial genome.
ACCESSION AF457060
VERSION AF457060.1 GI:22596302
KEYWORDS Human immunodeficiency virus 1 (HIV-1)
SOURCE Human immunodeficiency virus 1
ORGANISM Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
Lentivirus group.
REFERENCE 1 (bases 1 to 8806)
AUTHORS Dowling,W.E., Kim,B., Mason,C.J., Wasunna,K.M., Alam,U., Elson,L.,
Birx,D.L., Robb,M.L., McCutchan,F.E. and Carr,J.K.
TITLE Forty-one near full-length HIV-1 sequences from Kenya reveal an
epidemic of subtype A and A-containing recombinants
JOURNAL AIDS 16 (13), 1809-1820 (2002)
MEDLINE 22205599
PUBMED 12218394
REFERENCE 2 (bases 1 to 8806)
AUTHORS Dowling,W.E., Kim,B., Mason,C.J., Wasunna,K.Monique., Alam,U.,
Elson,L., Birx,D.L., Robb,M.L., McCutchan,F.E. and Carr,J.K.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-2001) Global Molecular Epidemiology, US Military
HIV Research Program - Walter Reed Army Institute of Research, 1
Taft Court, Rockville, MD 20850, USA
FEATURES
source
1. .8806
/organism="Human immunodeficiency virus 1"
/proviral
/mol_type="genomic DNA"
```

```
/isolate="00KE_KISII5009"
/db_xref="taxon:11676"
/country="Kenya"
/notes="subtype: A2/D recombinant"
1. .1467
/misc_feature /note="nonfunctional gag protein due to hypermutation"
1278. .4271
/misc_feature /note="nonfunctional pol protein due to hypermutation"
4216. .4794
/misc_feature /note="nonfunctional vif protein due to hypermutation"
4734. .5021
/gene="vpr"
4734. .5021
/gene="vpr"
/codon_start=1
/product="vpr protein"
/protein_id="AA03087.1"
/db_xref="GI:22596303"
/translation="MEQAPEDQGPQREPYNEWALELLEELKNEAVRHFPRMVLHSLGO
HIYETYGDTWAGVEATIRILQLLFIHFRIGCCQHSRIGILRRRANGPSRS"
5002. .7840
/gene="tat"
join(5002. .5216,7545. .7635)
/gene="tat"
/codon_start=1
/product="tat protein"
/protein_id="AA03088.1"
/db_xref="GI:22596304"
/translation="MDPVPNLEPNHPSQPTGCKNCYKCCYKCCFCFLNKGLG
ISYGRKRRPRGPPSPSSADHQNPVSKQSLPQTRISTGPESKKEVESKAETDRFA"
join(5141. .5216,7545. .7840)
/misc_feature /note="nonfunctional rev protein due to hypermutation"
5243. .5488
/gene="vpu"
5243. .5488
/gene="vpu"
/codon_start=1
/product="vpu protein"
/protein_id="AA03089.1"
/db_xref="GI:22596305"
/translation="MSSLOIVSIVGLIVAFILAVVTIVYIYRIKQKOKKIDWLK.
RISRAEDSGNESEGTTELSKLVEMGNLRFDLANDL"
5406. .7982
/misc_feature /note="nonfunctional envelope glycoprotein due to
hypermutation"
7984. .8622
/gene="nef"
7984. .8619
/gene="nef"
/codon_start=1
/product="nef protein"
/protein_id="AA03090.1"
/db_xref="GI:22596306"
/translation="MGGKWSSTCIIVGWPVEVREMRTPAAEEAQAAGVAVSODLAA
RGAVTSNRTTNSACAWLKAQESGEVGPVPPQPLRPMYKAAFDLGLFFLKERGG
LDGLIWSQRQDILLWVYNTQGFPPDQWNTYEGVREPLTEGRCFKLVPYDPSEIE
KATEGENSLHHPMCOHGIEDSEGEVLRMKWSSSLARRHWARELHPEYTKN"
8622. .8806
LTR /note="contains U3 and R"
BASE COUNT 3298 a 1561 c 1993 g 1954 t
ORIGIN
```

Alignment Scores:  
Pred. No.: 76.3 Length: 8806  
Score: 41.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-994-617-4 (1-9) x AF457060 (1-8806)

Qy 1 GlnlelleGlulysLeulleGlulys 9

Db 3282 CAAATAATAGAAAAAGTTAATAGAAAAAG 3308  
|||||

Search completed: August 6, 2003, 12:15:38  
Job time : 2098 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 6, 2003, 05:56:19 ; Search time 449 Seconds  
(without alignments)  
54,109 Million cell updates/sec

Title: US-09-994-617-4

Perfect score: 41

Sequence: 1 QIEKLIK 9

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_p2n.model -DEV=xl  
-Q/cgn2\_1/USPTO\_spool/US09994617/runat\_29072003\_160422\_7624/app\_query.fasta\_1.199  
-DB=N\_Geneseq\_19jun03 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTEXT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09994617 -CGCN\_1\_1\_471@runat\_29072003\_160422\_7624 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGUEURY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	100.0	9060	21	AAF57928 HIV-1 non-subtype
2	39	95.1	4851	21	AAAL13176 Vitellogenin encod
3	37	90.2	5703	24	ABK36429 HIV Cassette A1 DN
4	37	90.2	5703	24	ABK36432 HIV Cassette A2 DN
5	37	90.2	8509	22	AAI67035 Nucleotide sequenc
6	37	90.2	8509	24	AAI18307 DNA sequence of Ke
7	37	90.2	17373	24	ABK36435 HIV complete Savin
8	36	87.8	1807	13	AAQ20243 Mycoplasma-Specifi
9	36	87.8	5946	13	AAQ28398 Rat nestin gene.
10	36	87.8	11236	15	AAQ70447 Rat nestin gene -
11	36	87.8	910715	20	AAQ20248 Borrelia burgdorfe
c 12	36	87.8	1664976	19	AAV21209 Methanococcus jann
c 13	35	85.4	517	22	AAH34301 Human colon cancer
14	35	85.4	619	21	AAAL16220 Human colon cancer
15	35	85.4	1029	21	ABT23222 Human protein modi
16	35	85.4	1051	21	AACT7647 Human cancer assoc
17	35	85.4	1125	24	ABZ35293 Human gene express
18	35	85.4	1361	23	ABL05669 Drosophila melanog
19	35	85.4	3483	23	ABL05668 Drosophila melanog
20	35	85.4	3753	23	ABL05306 Drosophila melanog
21	35	85.4	4663	22	AAI58447 Human polynucleoti
22	35	85.4	640681	24	ABA92787 Buchnera sp. genom
23	34	82.9	60	24	ABN32659 Human spliced tran
24	34	82.9	274	25	ABX24714 Human GDP-mannose
25	34	82.9	439	25	ABZ19373 Group III CDNA can
26	34	82.9	474	25	ABZ19027 Group III CDNA can
27	34	82.9	498	25	ABZ19063 Group III CDNA can
28	34	82.9	661	18	AAT91127 Strawberry fruit a
29	34	82.9	1234	9	AAAN80494 Sequence encoding
30	34	82.9	1589	21	AAC44854 Arabidopsis thalia
31	34	82.9	1614	21	AAZ56974 Arabidopsis aldehy
32	34	82.9	2046	22	AAF57040 DNA sequence of P.
33	34	82.9	3150	22	AAF57038 DNA sequence of P.
34	34	82.9	3591	23	ABL07008 Drosophila melanog
c 35	34	82.9	4035	25	ABZ24687 Human immune syste
c 36	34	82.9	6265	24	ABL33417 Human immune syste
37	34	82.9	7116	23	ABL19292 Drosophila melanog
38	34	82.9	7921	24	ABL33971 Human immune syste
39	34	82.9	9229	9	AAAN80437 Entire sequence of
40	34	82.9	9229	19	AAV63467 HIV isolate LAV.MA
c 41	34	82.9	11670	24	ABL54326 Chemically treated
42	34	82.9	13827	23	ABL06940 Drosophila melanog
c 43	34	82.9	40324	24	ABO67150 Human angiogenesis
c 44	34	82.9	1830121	17	AAAT42063 Haemophilus influe
45	33	80.5	57	14	AAQ50237 HIV pol INS mutage

# ALIGNMENTS

RESULT 1

AAF57928

ID AAF57928 standard; DNA; 9060 BP.

XX

AC AAF57928;

XX

DT 20-APR-2001 (first entry)

XX

DE HIV-1 non-subtype B clone 94CV017-41 genome sequence.

XX

KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env;

KW vpu; vif; vpr; tat; rev; nef; vaccine; ds.

XX

OS Human immunodeficiency virus type 1.

XX

PN WO200026416-A1.

XX

PD 11-MAY-2000.  
 XX  
 PF 25-OCT-1999; 99WO-US24837.  
 PR 02-NOV-1998; 98US-0184418.  
 XX (UABR-) UAB RES FOUND.  
 XX Hahn BH, Shaw GM, Gao F;  
 XX WPI; 2000-365651/31.  
 DR  
 XX Novel genomic nucleic acids of non-subtype B human immunodeficiency  
 PT virus type 1 useful for detecting and treating AIDS comprises a  
 PT specific nucleotide sequence -  
 XX  
 XX Claim 5; Fig 13; 131pp; English.  
 XX  
 CC The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection.  
 XX  
 SQ Sequence 9060 BP; 3273 A; 1617 C; 2166 G; 2004 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 207 Length: 9060  
 Score: 41.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0  
 XX  
 US-09-994-617-4 (1-9) x AAF57928 (1-9060)  
 QY 1 GlnIleIleGluLysLeuIleGluLys 9  
 DB 3464 CAAATATAGAAAATTAATAGAAAG 3490  
 RESULT 2  
 ID AAAL3176 standard; DNA; 4851 BP.  
 XX AAAL3176;  
 AC  
 XX 20-JUL-2000 (first entry)  
 DT  
 XX Vitellogenin encoding gene vit-1 nucleotide sequence.  
 DE  
 XX Vitellogenin; vit-1; primer set; detect; endocrine disturbance;  
 KW toxicity; environmental contamination; ds.  
 XX  
 OS Caenorhabditis elegans.  
 XX JP2000069977-A.  
 PN  
 XX 07-MAR-2000.  
 PD  
 XX 03-SEP-1998; 98JP-0249723.  
 PF  
 XX 03-SEP-1998; 98JP-0249723.  
 PR  
 XX (KUMA-) KUMAMOTO KOTAI KENKYUSHO KK.  
 PA  
 XX WPI; 2000-264452/23.  
 DR  
 XX A primer for the amplification of nematode vitellogenin gene - for  
 PT evaluation of the endocrine disturbing activity of a chemical substance  
 PT  
 XX

PS Claim 3; Page 8-10; 17pp; Japanese.  
 XX  
 CC This sequence represents the Caenorhabditis elegans vit-1 gene, which  
 CC encodes vitellogenin. The present invention relates to a primer set (see  
 CC AAAL3179-A13188) containing primers which hybridise to different parts  
 CC of the common regions of the vit-1, vit-2, vit-3, vit-4 and vit-5 genes.  
 CC The vit-1 to 5 genes encode nematode vitellogenin. The invention also  
 CC includes a method for the detection of a vitellogenin gene in which a  
 CC gene encoding nematode vitellogenin is amplified using the primer set. A  
 CC method for evaluating the endocrine disturbing activity of a chemical  
 CC substance in which a gene encoding vitellogenin is amplified using the  
 CC primer set is also included in the invention. The method for evaluation  
 CC can be used for research on the toxicity of chemical substances and  
 CC environmental contamination.  
 XX  
 SQ Sequence 4851 BP; 1391 A; 1350 C; 999 G; 1111 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 272 Length: 4851  
 Score: 39.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 95.12% Indels: 0  
 DB: 21 Gaps: 0  
 XX  
 US-09-994-617-4 (1-9) x AAAL3176 (1-4851)  
 QY 1 GlnIleIleGluLysLeuIleGluLys 9  
 DB 2371 CAACTCATTGAGAAAGCTCATTGAGAAG 2397  
 RESULT 3  
 ID ABK36429 standard; DNA; 5703 BP.  
 XX ABK36429;  
 AC  
 XX 08-MAY-2002 (first entry)  
 DT  
 XX HIV Cassette A1 DNA.  
 DE  
 XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia; ds.  
 XX  
 OS Human immunodeficiency virus type 1.  
 OS Synthetic.  
 XX WO200190197-A1.  
 PN  
 XX 29-NOV-2001.  
 PD  
 XX 25-MAY-2001; 2001WO-AU00622.  
 PF  
 XX 26-MAY-2000; 2000AU-0007761.  
 PR  
 XX (AUSU) UNIV AUSTRALIAN NAT.  
 PA  
 XX Thomson SA, Ramshaw IA;  
 PI  
 XX WPI; 2002-147575/19.  
 DR  
 XX P-PSDB; AAU84590.  
 DR  
 XX New synthetic polypeptides having several different segments of at  
 PT least one parent polypeptide linked together differently compared to  
 PT the linkage in the parent polypeptide, for inducing immune response  
 PT against a pathogen or cancer -  
 XX  
 XX Example 1; Fig 30; 364pp; English.  
 PS  
 XX The invention relates to a new synthetic polypeptide (I) comprising  
 CC several different segments of at least one parent polypeptide linked

CC together in a different relationship relative to their linkage in the  
 CC parent polypeptide to impede, abrogate or otherwise alter at least one  
 CC function associated with the parent polypeptide and for inducing an  
 CC immune response against a pathogen or cancer. Also included are a  
 CC synthetic polynucleotide encoding and a computer system for  
 CC designing the synthetic polypeptides. The synthetic polypeptides and  
 CC polynucleotides are referred to as a Savine. The synthetic polypeptide is  
 CC useful for modulating immune responses preferably directed against a  
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,  
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone  
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.  
 CC Compositions comprising the polypeptide may be used in the treatment or  
 CC prophylaxis against viral (such as infections caused by HIV (human  
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis  
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial  
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,  
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic  
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,  
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present sequence  
 CC encodes a cassette protein consisting of several peptides derived from  
 CC a parent protein. One or more cassettes are used to construct a savine of  
 CC the invention.

SQ Sequence 5703 BP; 1708 A; 1478 C; 1407 G; 1110 T; 0 other;

Alignment Scores:  
 Pred. No.: 868 Length: 5703  
 Score: 37.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 90.24% Indels: 0  
 DB: 24 Gaps: 0

US-09-994-617-4 (1-9) x ABK36429 (1-5703)

Qy 1 GlnlleleGluLysLeulleGluLys 9

Db 1729 CAGATTATCGAAAGCTCATCAAAAAG 1755

RESULT 4

ABK36432

ID ABK36432 standard; DNA; 5703 BP.

AC ABK36432;

DT 08-MAY-2002 (first entry)

XX HIV Cassette A2 DNA.

XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia; ds.

XX Human immunodeficiency virus type 1.

OS Synthetic.

PN WO200190197-A1.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-AU00622.

XX 26-MAY-2000; 2000AU-0007761.

XX (AUSU) UNIV AUSTRALIAN NAT.

XX Thomson SA, Ramshaw IA;

XX WPI; 2002-147575/19.

DR P-PSDB; AA084593.

XX New synthetic polypeptides having several different segments of at

PT least one parent polypeptide linked together differently compared to  
 PT the linkage in the parent polypeptide, for inducing immune response  
 XX against a pathogen or cancer

PS Example 1; Flg 30; 364pp; English.

XX The invention relates to a new synthetic polypeptide (I) comprising  
 CC several different segments of at least one parent polypeptide linked  
 CC together in a different relationship relative to their linkage in the  
 CC parent polypeptide to impede, abrogate or otherwise alter at least one  
 CC function associated with the parent polypeptide and for inducing an  
 CC immune response against a pathogen or cancer. Also included are a  
 CC synthetic polynucleotide encoding and a computer system for  
 CC designing the synthetic polypeptides. The synthetic polypeptides and  
 CC polynucleotides are referred to as a Savine. The synthetic polypeptide is  
 CC useful for modulating immune responses preferably directed against a  
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,  
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone  
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.  
 CC Compositions comprising the polypeptide may be used in the treatment or  
 CC prophylaxis against viral (such as infections caused by HIV (human  
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis  
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial  
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,  
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic  
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,  
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present sequence  
 CC encodes a cassette protein consisting of several peptides derived from  
 CC a parent protein. One or more cassettes are used to construct a savine of  
 CC the invention.

SQ Sequence 5703 BP; 1718 A; 1488 C; 1397 G; 1100 T; 0 other;

Alignment Scores:

Pred. No.: 868 Length: 5703  
 Score: 37.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 90.24% Indels: 0  
 DB: 24 Gaps: 0

US-09-994-617-4 (1-9) x ABK36432 (1-5703)

Qy 1 GlnlleleGluLysLeulleGluLys 9

Db 1729 CAGATTATCGAAAGCTTATCAAAAAG 1755

RESULT 5

AAI67035

ID AAI67035 standard; DNA; 8509 BP.

XX AAI67035;

DT 11-FEB-2002 (first entry)

XX Nucleotide sequence of an endogenous ketogulonigenium plasmid.

XX Ketogulonigenium; plasmid; strain NRRL B-30035; vector; ds.

OS Synthetic.

PN WO200177159-A2.

XX 18-OCT-2001.

XX 05-APR-2001; 2001WO-US11058.

XX 05-APR-2000; 2000US-194624P.

XX (UNMS) UNIV MICHIGAN STATE.

PA (ARCH) ARCHER-DANIELS MIDLAND CO.

PI Schmidt TM, Stoddard SF;

XX WPI; 2001-657165/75.  
 XX New nucleic acid comprising the sequence of a Ketogulonigenium plasmid  
 PT designated pADM291 is endogenous to microorganism strain NRRL B-30035  
 PT  
 XX Claim 1; Fig 1A-E; 14pp; English.  
 XX  
 CC The present sequence represents a purified nucleic acid of a  
 CC Ketogulonigenium plasmid endogenous to strain NRRL B-30035. The  
 CC polynucleotide can be used in a vector. The vector comprises either  
 CC (i) the polynucleotide and a marker gene or (ii) the polynucleotide, a  
 CC promoter, a transcription terminator and a discrete series of restriction  
 CC endonuclease recognition sites between the promoter and transcription  
 CC terminator.  
 XX  
 SQ Sequence 8509 BP; 2065 A; 2239 C; 2345 G; 1860 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.36e+03 Length: 8509  
 Score: 37.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 90.24% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-994-617-4 (1-9) x AAI67035 (1-8509)  
 QY 1 GlnlleleclulysleulleGluLys 9  
 DB 2650 CAAATCTGGAAAAATGCTTGAAGAAG 2676  
 RESULT 6  
 AAS18307  
 ID AAS18307 standard; DNA; 8509 BP.  
 XX  
 AC AAS18307;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE DNA sequence of Ketogulonigenium endogenous plasmid pADM291.  
 XX  
 KW Cloning vector; Ketogulonigenium replicon; endogenous plasmid;  
 KW transformed host cell; Escherichia coli; pADM291; circular; cyclic; ds.  
 XX  
 OS Ketogulonigenium sp. strain ADM291-19.  
 XX  
 PN WO200177347-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 XX 05-APR-2001; 2001WO-US11059.  
 XX  
 XX 05-APR-2000; 2000US-194625P.  
 XX  
 PA (ARCH ) ARCHER-DANIELS MIDLAND CO.  
 PA (DELI/) D'ELIA J.  
 XX  
 XX D'Elia J;  
 XX  
 DR WPI; 2002-049150/06.  
 XX  
 XX Novel nucleic acid vector comprising Ketogulonigenium replicon found on  
 PT a specific deposited endogenous plasmid, useful for producing  
 PT polypeptides and/or transcripts by culturing host cells transformed  
 PT with vector  
 XX  
 PS Example 7; Fig 2; 66pp; English.  
 XX  
 CC The present invention relates to the isolation of vectors comprising  
 CC a Ketogulonigenium replicon found on the endogenous plasmid, pADM291.  
 CC The invention also describes methods of transforming host cells with

CC the vectors and producing polypeptides and/or antisense transcripts by  
 CC culturing the transformed host cells. The vectors are useful for  
 CC transforming a host cell by conjugation or electroporation.  
 CC The vectors which have a replicon functional in both Ketogulonigenium  
 CC and Escherichia coli, enable the cloning of certain genes of  
 CC Ketogulonigenium in E.coli as the latter is an efficient host for  
 CC amplification of vector DNA. The present DNA sequence represents the  
 CC Ketogulonigenium endogenous plasmid pADM291.  
 XX  
 SQ Sequence 8509 BP; 2065 A; 2239 C; 2345 G; 1860 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.36e+03 Length: 8509  
 Score: 37.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 90.24% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-994-617-4 (1-9) x AAS18307 (1-8509)  
 QY 1 GlnlleleclulysleulleGluLys 9  
 DB 2650 CAAATCTGGAAAAATGCTTGAAGAAG 2676  
 RESULT 7  
 ABK36435  
 ID ABK36435 standard; DNA; 17373 BP.  
 XX  
 AC ABK36435;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE HIV complete Savine DNA.  
 XX  
 KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;  
 KW viral infection; human immunodeficiency virus; melanoma;  
 KW bacterial infection; Salmonella; Legionella; parasitic infection;  
 KW Trypanosoma; Toxoplasma; Giardia; ds.  
 XX  
 OS Human immunodeficiency virus type 1.  
 OS Synthetic.  
 XX  
 PN WO200190197-Al.  
 XX  
 XX 29-NOV-2001.  
 XX  
 XX 25-MAY-2001; 2001WO-AU00622.  
 XX  
 XX 26-MAY-2000; 2000AU-0007761.  
 XX  
 PA (AUSU ) UNIV AUSTRALIAN NAT.  
 XX  
 PI Thomson SA, Ramshaw IA;  
 XX  
 DR WPI; 2002-147575/19.  
 DR P-PSDB; AAU84596.  
 XX  
 XX New synthetic polypeptides having several different segments of at  
 PT least one parent polypeptide linked together differently compared to  
 PT the linkage in the parent polypeptide, for inducing immune response  
 PT against a pathogen or cancer  
 XX  
 XX Example 1; Fig 15; 364pp; English.  
 PS  
 PS The invention relates to a new synthetic polypeptide (I) comprising  
 CC several different segments of at least one parent polypeptide linked  
 CC together in a different relationship relative to their linkage in the  
 CC parent polypeptide to impede, abrogate or otherwise alter at least one  
 CC function associated with the parent polypeptide and for inducing an  
 CC immune response against a pathogen or cancer. Also included are a  
 CC synthetic polynucleotide encoding and a computer system for  
 CC designing the synthetic polypeptides. The synthetic polypeptides and



CC polynucleotides are referred to as a Savine. The synthetic polypeptide is  
 CC useful for modulating immune responses preferably directed against a  
 CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,  
 CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone  
 CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.  
 CC Compositions comprising the polypeptide may be used in the treatment or  
 CC prophylaxis against viral (such as infections caused by HIV (human  
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis  
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial  
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,  
 CC Salmonella, Streptococcal, Legionella and Mycobacterium) or parasitic  
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,  
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present  
 CC sequence encodes a savine protein of the invention.

XX SQ Sequence 17373 BP; 5020 A; 4501 C; 4384 G; 3375 T; 4 U; 89 other;

#### Alignment Scores:

Pred. No.: 3.05e+03 Length: 17373  
 Score: 37.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 90.24% Indels: 0  
 DB: 24 Gaps: 0

US-09-994-617-4 (1-9) x ABK36435 (1-17373)

Qy 1 GlnllellegluLysLeuileglulys 9  
 Db 1765 CAGATTATCGAAAGCTCATCAAAAG 1791

#### RESULT 8

AAQ20243  
 ID AAQ20243 standard; DNA; 1807 BP.

XX AC AAQ20243;  
 XX DT 25-MAR-2003 (updated)  
 XX DT 25-MAR-1992 (first entry)  
 XX DE Mycoplasma-specific DNA fragment.

XX KW Detection; ss.

XX OS Mycoplasma hyopneumoniae.

XX PN JP03254700-A.

XX PD 13-NOV-1991.

XX PF 06-MAR-1990; 90JP-0052687.

XX PR 06-MAR-1990; 90JP-0052687.

XX PA (TAKI) TAKARA SHUZO CO LTD.

XX XX WPI; 1992-002469/01.

XX PT Highly sensitive detection of mycoplasma - by detecting region of  
 PT restriction enzyme map and amplifying DNA sequence

XX PS Disclosure; Fig 3; 10pp; Japanese.

XX CC The sequence is a 1807bp fragment of M. hyopneumoniae and was used  
 CC to test a kit for the detection of pig mycoplasma. The DNA is first  
 CC amplified by PCR and then the probe is used to detect the amplified  
 CC DNA. This method is highly sensitive. See also AAQ20243-Q20245 and  
 CC AAQ21024-Q21027.

XX CC (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 1807 BP; 698 A; 205 C; 254 G; 648 T; 2 other;

#### Alignment Scores:

Pred. No.: 386 Length: 1807  
 Score: 36.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 87.80% Indels: 0  
 DB: 13 Gaps: 0

US-09-994-617-4 (1-9) x AAQ20243 (1-1807)

Qy 1 GlnllellegluLysLeuileglulys 9

Db 1461 GAATTATGAAACCTTATTGAAAG 1487

#### RESULT 9

AAQ28398  
 ID AAQ28398 standard; DNA; 5946 BP.

XX AC AAQ28398;

XX DT 25-MAR-2003 (updated)

XX DT 15-FEB-1993 (first entry)

XX DE Rat nestin gene.

XX KW Intermediate filament; central nervous system; brain tumour;  
 KW neurofilament; ss.

XX OS Rattus norvegicus.

XX FH Key Location/Qualifiers

FT CDS 128..5545

FT FT /\*tag= a

FT FT /product= nestin

FT FT /note= "128 could be 161"

FT FT 1..913

FT FT /\*tag= b

FT FT /number= 1

FT FT 914..1038

FT FT /\*tag= c

FT FT /number= 2

FT FT 1039..1110

FT FT /\*tag= d

FT FT /number= 3

FT FT 1111..5946

FT FT /\*tag= e

FT FT /number= 4

FT FT 5924..5929

FT FT /\*tag= f

FT FT polyA\_signal

XX PN WO9214821-A2.

XX PD 03-SEP-1992.

XX PF 21-FEB-1992; 92WO-US01375.

XX PR 22-FEB-1991; 91US-0660412.

XX XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX PI Lendahl U, McKay RDG;

XX DR WPI; 1992-316175/38.

XX DR P-PSDB; AAR27204.

XX PT Diagnosis of pre-disposition to brain tumours - using DNA  
 PT encoding nestin which distinguishes neural multi-potential stem  
 PT cells from neuronal, glial and muscle cells.

XX PS Claim 3; Fig 1; 63pp; English.

XX CC A lambda gtlII cDNA expression library was constructed from poly(A)+  
 CC RNA from CNS of day 15 rat embryos. The library was screened with  
 CC MAb Rat 401. cDNA from the most immunopositive clone was used as

CC a probe to isolate additional clones from a lambda gt10 library  
 CC from rat E15 CNS. Hybridising clones were sequenced; the remainder  
 CC of the nestin gene was obtained from genomic sequences. There are  
 CC two potential initiation codons at positions 128 and 160. Sequence  
 CC comparison of cDNA and genomic clones reveals three introns at  
 CC positions 912, 1038 and 1111. The sequence has significant  
 CC similarities to the five classes of intermediate filaments.  
 CC The coding sequence can be used to develop probes for  
 CC detecting predisposition to brain tumours; the expression of nestin  
 CC distinguishes neural multi potential stem cells and brain tumour  
 CC cells from the more differentiated neural cell types (e.g. neuronal,  
 CC glial and muscle cells) of the mammalian brain.  
 CC NOTE: the rat nestin coding sequence is printed in the patent  
 CC specification but is illegible; this sequence has been obtained from  
 CC the EMBL record (Accession number M34394).  
 CC See also Q28399 for the human nestin gene.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 5946 BP; 1600 A; 1401 C; 1871 G; 1074 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.48e+03 Length: 5946  
 Score: 36.00 Matches: 7  
 Percent Similarity: 100.00% Conservatives: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 87.80% Indels: 0  
 DB: 13 Gaps: 0

US-09-994-617-4 (1-9) x AAQ28398 (1-5946)

QY 1 Glnlleclglulysleulleglulys 9  
 |||||:||||:|||||  
 Db 2264 CAGATCTTGAGAGACTGATAGAAAAA 2290

RESULT 10

AAQ70447

ID AAQ70447 standard; DNA; 11236 BP.

XX AAQ70447;

DT 25-MAR-2003 (updated)

DT 21-MAR-1995 (first entry)

XX

DE Rat nestin gene - its product is useful to identify brain tumours.  
 XX nestin gene; brain tumour; neoplastic cells; glial; neuronal;

KW muscle; neural multipotential stem cell; mammalian brain; detection;  
 KW diagnosis; medulloblastoma; glioblastoma; oligodendroglioma; ds.  
 XX

OS Rattus rattus.

XX Key Location/Qualifiers  
 FH misc\_feature 2463

FT /tag= a  
 FT /note= "start of primary transcript"  
 FT 2389..10821

FT /tag= b

FT /product= Nestin protein  
 FT 3089..3090

FT /tag= c

FT /transl\_except= pos:3087..3091, aa:His, Arg  
 FT /note= "sequence should be CAY CGC, ie. Y or C has  
 FT been deleted in the sequence given"

FT misc\_feature 3104

FT /tag= d  
 FT /note= "apparent inclusion of a nucleotide"  
 FT 3375..4339

FT intron

FT /tag= e

FT 4340..4464

FT /tag= f

FT 4465..6119

FT /tag= g

FT 6120..6193

FT exon

FT intron  
 FT 6194..6388  
 FT /tag= i  
 FT 6389..10821  
 FT /tag= j  
 XX  
 PN US5338839-A.  
 XX  
 PD 16-AUG-1994.  
 XX  
 PF 19-MAR-1992; 92US-0853913.  
 XX  
 PR 12-APR-1988; 88US-0180548.  
 PR 02-JUN-1988; 88US-0201762.  
 PR 25-OCT-1990; 90US-0603803.  
 PR 22-FEB-1991; 91US-0660412.  
 PR 19-MAR-1992; 92US-0853913.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Lendahl U, McKay RDG;  
 XX  
 DR WPI; 1994-263332/32.  
 DR P-PSDB; AAR60126.  
 XX  
 PT Nucleotide and protein sequences for human and rat nestin -  
 PT distinguishes neural multipotential stem cells and brain tumour  
 PT cells from more differentiated cell types; for use in the  
 PT diagnosis of brain tumours

Claim 1; Column 25-34; 45pp; English.

PS AAQ70447 is the rat nestin gene encoding nestin protein (AAR60126).  
 CC Nestin protein expression distinguishes neural multipotential stem  
 CC cells and brain tumour cells from the more differentiated neural  
 CC cell types (eg., neuronal, glial and muscle cells of the adult brain).  
 CC The nestin protein can be used in diagnosing tumours of the brain,  
 CC such as medulloblastomas, glioblastomas and oligodendroglioma.  
 CC (See also AAQ70448).  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 11236 BP; 2876 A; 2678 C; 3258 G; 2424 T; 0 other;

Alignment Scores:

Pred. No.: 3.04e+03 Length: 11236  
 Score: 36.00 Matches: 7  
 Percent Similarity: 100.00% Conservatives: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 87.80% Indels: 0  
 DB: 15 Gaps: 0

US-09-994-617-4 (1-9) x AAQ70447 (1-11236)

QY 1 Glnlleclglulysleulleglulys 9  
 |||||:||||:|||||  
 Db 7540 CAGATCTTGAGAGACTGATAGAAAAA 7566

RESULT 11

AAQ20248

ID AAX20248 standard; DNA; 910715 BP.

XX AAX20248;

XX 04-MAY-1999 (first entry)

DT

DE Borrelia burgdorferi polynucleotide sequence #1.

XX

KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 XX

XX Borrelia burgdorferi.

XX

PN WO9858943-A1.  
 XX 30-DEC-1998.  
 XX 18-JUN-1998; 98WO-US12764.  
 XX 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMMUNE INC.  
 XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;  
 PI White OR;  
 DR WPI; 1999-081217/07.  
 XX New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease  
 XX  
 PS Claim 1; Page 157-671; 1128pp; English.  
 XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from  
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.  
 XX  
 SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;  
 Alignment Scores:  
 Pred. No.: 4.17e+05 Length: 910715  
 Score: 36.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 87.80% Indels: 0  
 DB: 20 Gaps: 0  
 US-09-994-617-4 (1-9) x AAX20248 (1-910715)  
 QY 1 GlnllellegluLysLeuileclLys 9  
 Db 815320 GAAATTTAGAAAAATTAATAGAAAA 815346  
 RESULT 12  
 AAV21209/c  
 ID AAV21209 standard; DNA; 1664976 BP.  
 XX AAV21209;  
 AC AAV21209;  
 XX 10-NOV-1998 (first entry)  
 DT  
 DE Methanococcus jannaschii circular chromosome.  
 DE Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 KW genome; autotrophic; extrachromosomal element; identification; ds.  
 KW Methanococcus jannaschii.  
 OS  
 XX WO9807830-A2.  
 PN  
 XX 26-FEB-1998.  
 PD  
 XX 22-AUG-1997; 97WO-US14900.  
 PF  
 XX 22-AUG-1996; 96US-0024428.  
 PR

XX (GENO-) INST GENOMIC RES.  
 PA (UNII) UNIV ILLINOIS FOUND.  
 PA (UWJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
 PI WPI; 1998-169145/15.  
 DR Complete genome sequence of methano-genic archaeon, *Methanococcus*  
 PT *jannaschii* - useful in identification of *M. jannaschii* genome  
 PT fragment  
 XX  
 PS Claim 13; Page 152-585; 614pp; English.  
 XX The present sequence represents the complete 1.66-megabase pair genome  
 CC sequence of the *Methanococcus jannaschii* circular chromosome. The  
 CC present invention describes *M. jannaschii* open reading frames from the  
 CC genome sequence. The invention also describes a computer based system  
 CC for identifying fragments of the *M. jannaschii* genome that are  
 CC homologous to target nucleotide sequences, comprising: (a) data storage  
 CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
 CC bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide  
 CC sequence at least 99.9% identical to it; (b) search means for comparing a  
 CC target sequence to the nucleotide sequence of the data storage means to  
 CC identify a homologous sequence, and (c) retrieval means for obtaining  
 CC the homologous sequence. The method, which is based on whole genome  
 CC random sequencing of an autotrophic archaeon *M. jannaschii*, the genome  
 CC of which consists of 3 physically distinct elements, a large circular  
 CC chromosome (the 1664976 bp sequence given in AAV21209), a large circular  
 CC extra-chromosomal element (the 58407 bp sequence given in AAV21210), and  
 CC a small circular extra-chromosomal element (the 16550 bp sequence given  
 CC in AAV21211), can be used in the identification of *M. jannaschii* genome  
 CC fragment.  
 XX  
 SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;  
 Alignment Scores:  
 Pred. No.: 7.92e+05 Length: 1664976  
 Score: 36.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 87.80% Indels: 0  
 DB: 19 Gaps: 0  
 US-09-994-617-4 (1-9) x AAV21209 (1-1664976)  
 QY 1 GlnllellegluLysLeuileclLys 9  
 Db 1054677 GAGATTCGGAATAATGATAGAAAAG 1054651  
 RESULT 13  
 AAH34301/c  
 ID AAH34301 standard; cDNA; 517 BP.  
 XX AAH34301;  
 AC AAH34301;  
 XX 03-SEP-2001 (first entry)  
 DT  
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1383.  
 DE Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma; ss.  
 KW Homo sapiens.  
 OS  
 XX WO200122920-A2.  
 PN  
 XX 05-APR-2001.  
 PD  
 XX 28-SEP-2000; 2000WO-US26524.  
 PF  
 XX 29-SEP-1999; 99US-0157137.  
 PR

```
PR 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR P-PSDB; AAG74896.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT
XX Claim 1; Page 3093; 9803pp; English.
PS
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX SQ Sequence 517 BP; 188 A; 85 C; 77 G; 163 T; 4 other;

Alignment Scores:
Pred. No.: 153 Length: 517
Score: 35.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 85.37% Indels: 0
DB: 22 Gaps: 0

US-09-994-617-4 (1-9) x AAH34301 (1-517)
QY 1 GlnlleleclulysleulleGluLys 9
Db 455 GAAGTCTAGAGAAATATAGAAAAA 429

RESULT 14
AAH16220
ID AAH16220 standard; DNA; 619 BP.
XX
XX AC AAH16220;
XX
XX DT 14-JUN-2000 (first entry)
XX
XX DE Human colon cancer differentially expressed nucleotide sequence #225.
XX
XX KW Colon cancer; detect; differential expression; human; treatment;
XX KW detect mutation; non-invasive diagnostic method; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200012702-A2.
XX
XX PD 09-MAR-2000.
XX
XX PF 30-AUG-1999; 99WO-US19424.
XX
XX PR 31-AUG-1998; 98US-0098639.
XX
XX PR 27-JAN-1999; 99US-0117393.
XX

(FARB ) BAYER CORP.
Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
Schlegel R;
WPI; 2000-256641/22.
XX
XX Novel nucleic acids and proteins for identifying therapeutic agents
PT useful for treating and diagnosing cancer, especially colon cancer -
PT
XX Claim 16; Page 213-214; 345pp; English.
PS
XX This sequence represents a human nucleotide sequence which is
CC differentially expressed in colon cancer cells compared to the expression
CC levels in normal cells. The nucleotide sequence can be used as a source
CC of primers and probes. The nucleotide sequence is useful for determining
CC the phenotype of a cell by detecting the differential expression of the
CC sequence relative to a normal cell. The probes derived from the sequence
CC can also be used to determine the phenotype of cells in a sample. Probes
CC and antibodies which hybridize to the nucleotide sequence can also be
CC used to determine the phenotype of a cell. The primers are useful for
CC detecting a mutation in a test nucleotide sequence and also for detecting
CC cancer, preferably colon cancer. Antibodies against the protein encoded
CC by the nucleotide sequence can also be used in a method to detect colon
CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
CC colon cancer at an early stage.
XX
XX SQ Sequence 619 BP; 186 A; 121 C; 151 G; 156 T; 5 other;

Alignment Scores:
Pred. No.: 188 Length: 619
Score: 35.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 85.37% Indels: 0
DB: 21 Gaps: 0

US-09-994-617-4 (1-9) x AAH16220 (1-619)
QY 1 GlnlleleclulysleulleGluLys 9
Db 501 GAAGTCTAGAGAAATATAGAAAAA 527

RESULT 15
ABT23222
ID ABT23222 standard; DNA; 1029 BP.
XX
XX AC ABT23222;
XX
XX DT 01-MAY-2003 (first entry)
XX
XX DE Human protein modification + maintenance molecule DNA SEQ ID No 51.
XX
XX KW Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;
XX KW cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;
XX KW antiulcer; hepatotropic; gynaecological; antibacterial; virucide;
XX KW protozoacide; antiparasitic; cell proliferative disease; PMOD;
XX KW protein modification and maintenance molecule; immunogenic fragment;
XX KW cancer; autoimmune; inflammatory disease; neurological disorder;
XX KW gastrointestinal; developmental; vesicle trafficking disorder; infection;
XX KW protein-protein interaction; drug-target interaction;
XX KW gene expression profile; human; gene; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO2003000844-A2.
XX
XX PD 03-JAN-2003.
XX
XX PF 18-JUN-2002; 2002WO-US19360.
XX
```

Search completed: August 6, 2003, 11:41:37  
Job time : 616 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: August 6, 2003, 11:31:47 ; Search time 118 Seconds  
(without alignments)  
33.665 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIIEKLEK 9

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delopt 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US0994617/runat\_29072003\_160423\_7666/app\_query.fasta\_1.199  
-DB=issued Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09994617@cgn2\_1.1.85 @runat\_29072003\_160423\_7666 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	3012	4	US-09-184-418C-92
2	41	100.0	9060	4	US-09-184-418C-92
3	37	90.2	1482	4	US-09-328-352-2627
4	37	90.2	8509	4	US-09-826-205-1
5	36	87.8	11236	1	US-07-853-913-1
6	36	87.8	1664976	4	US-08-916-421B-1
7	35	85.4	619	3	US-09-385-982-225
8	35	85.4	4663	4	US-09-620-312D-328
9	34	82.9	1830121	4	US-09-557-884-1
10	34	82.9	1830121	4	US-09-643-990A-1
11	33	80.5	56	2	US-08-850-049-60
12	33	80.5	56	2	US-08-850-049-61

13	33	80.5	56	2	US-08-050-478-60	Sequence 60, Appl
14	33	80.5	56	2	US-08-050-478-61	Sequence 61, Appl
15	33	80.5	56	2	US-09-414-117-60	Sequence 60, Appl
16	33	80.5	56	3	US-09-414-117-61	Sequence 61, Appl
17	33	80.5	56	4	US-09-678-437-60	Sequence 60, Appl
18	33	80.5	56	4	US-09-678-437-61	Sequence 61, Appl
c 19	33	80.5	431	4	US-09-222-575-115	Sequence 115, App
c 20	33	80.5	431	4	US-09-389-681-115	Sequence 115, App
c 21	33	80.5	431	4	US-09-620-405B-115	Sequence 115, App
c 22	33	80.5	431	4	US-09-339-338-115	Sequence 115, App
c 23	33	80.5	431	4	US-09-433-826B-115	Sequence 115, App
c 24	33	80.5	431	4	US-09-604-287A-115	Sequence 115, App
c 25	33	80.5	1680	5	PCT-US93-06748-1	Sequence 1, Appl
26	33	80.5	2190	4	US-09-107-532A-316	Sequence 316, App
27	33	80.5	2348	2	US-08-876-546A-1	Sequence 1, Appl
28	33	80.5	2348	3	US-09-412-252-1	Sequence 1, Appl
29	33	80.5	2348	4	US-09-079-675-1	Sequence 1, Appl
30	33	80.5	2601	3	US-09-117-217-7	Sequence 7, Appl
31	33	80.5	2601	3	US-09-117-217-9	Sequence 9, Appl
32	33	80.5	2601	3	US-09-117-217-11	Sequence 11, Appl
33	33	80.5	2601	3	US-09-117-217-13	Sequence 13, Appl
34	33	80.5	2601	4	US-09-735-487-7	Sequence 7, Appl
35	33	80.5	2601	4	US-09-735-487-9	Sequence 9, Appl
36	33	80.5	2601	4	US-09-735-487-11	Sequence 11, Appl
37	33	80.5	2601	4	US-09-735-487-13	Sequence 13, Appl
38	33	80.5	2739	2	US-07-743-357-16	Sequence 16, Appl
39	33	80.5	2739	4	US-09-184-418C-22	Sequence 22, Appl
40	33	80.5	3000	4	US-09-184-418C-74	Sequence 74, Appl
41	33	80.5	3012	4	US-09-184-418C-57	Sequence 57, Appl
42	33	80.5	3017	4	US-09-184-418C-83	Sequence 83, Appl
43	33	80.5	3033	2	US-07-743-357-19	Sequence 19, Appl
44	33	80.5	3045	4	US-09-319-588C-5	Sequence 5, Appl
45	33	80.5	3856	2	US-07-743-357-20	Sequence 20, Appl

#### ALIGNMENTS

RESULT 1  
US-09-184-418C-92  
; Sequence 92, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 92  
; LENGTH: 3012  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=94CY017.41; gene=pol  
US-09-184-418C-92

Alignment Scores:  
Pred. No.: 3.82 Length: 3012  
Score: 41.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-994-617-4 (1-9) x US-09-184-418C-92 (1-3012)

QY 1 GlnIleIleGluLysLeuIleGluLys 9

Db 2023 CAAATAAGATAAGAAATTAATAGAAAG 2049

```
RESULT 2
US-09-184-418C-10
; Sequence 10, Application US/09184418C
; Patent No. 6492110
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice
; APPLICANT: Gao, Feng
; APPLICANT: Shaw, George
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN
; FILE REFERENCE: D6287
; CURRENT APPLICATION NUMBER: US/09/184.418C
; CURRENT FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 112
; SEQ ID NO 10
; LENGTH: 9060
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: Isolate-94CY017.41; 159.1649:"gag";
; OTHER INFORMATION: 1442.4453:"pol"; 4398.4976:"vif"; 4916.5206:"vpr";
; OTHER INFORMATION: 5187.7841:"tat"; 5326.8046:"rev";
; OTHER INFORMATION: 5428.5673:"vpu"; 5591.8188:"env"; 8190.8843:"nef"
US-09-184-418C-10
Alignment Scores:
Pred. No.: 14.4 Length: 9060
Score: 41.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-09-994-617-4 (1-9) x US-09-184-418C-10 (1-9060)
OY 1 GlnlleleGluLysLeulleGluLys 9
Db 3464 CAAATAATAGAAAAATTAATAGAAAAG 3490
RESULT 3
US-09-328-352-2627
; Sequence 2627, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2627
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2627
Alignment Scores:
Pred. No.: 13.8 Length: 1482
Score: 37.00 Matches: 8
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 1
Query Match: 90.24% Indels: 0
DB: 4 Gaps: 0
US-09-994-617-4 (1-9) x US-09-328-352-2627 (1-1482)
OY 1 GlnlleleGluLysLeulleGluLys 9
Db 928 CAAATTTTGGAAAACTGATTGAAAAA 954
RESULT 4
US-09-826-205-1
```

```
; Sequence 1, Application US/09826205
; Patent No. 6503748
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Thomas, M.
; APPLICANT: Stoddard, Steven, F.
; TITLE OF INVENTION: An Endogenous Ketogulonigenium Plasmid
; FILE REFERENCE: 1533.0950001
; CURRENT APPLICATION NUMBER: US/09/826.205
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 8509
; TYPE: DNA
; ORGANISM: Ketogulonigenium
US-09-826-205-1
Alignment Scores:
Pred. No.: 113 Length: 8509
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 90.24% Indels: 0
DB: 4 Gaps: 0
US-09-994-617-4 (1-9) x US-09-826-205-1 (1-8509)
OY 1 GlnlleleGluLysLeulleGluLys 9
Db 2650 CAAATCTGGAAAAATGCTTGAAGAAG 2676
RESULT 5
US-07-853-913-1
; Sequence 1, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853.913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
```



```
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11236 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-853-913-1

Alignment Scores:
Pred. No.: 271 Length: 11236
Score: 36.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 77.80% Indels: 0
DB: 1 Gaps: 0

US-09-994-617-4 (1-9) x US-07-853-913-1 (1-11236)

QY 1 GlnllelleglLysLeuileglLys 9
|||||:|||||:|||||:|||||
Db 7540 CAGATACCTGAGAGACTGATAGAAAA 7566

RESULT 6
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916.421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024.428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
```

```
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

```
Alignment Scores:
Pred. No.: 1 07e+05 Length: 1664976
Score: 36.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 87.80% Indels: 0
DB: 4 Gaps: 0
```

US-09-994-617-4 (1-9) x US-08-916-421B-1 (1-1664976)

```
QY 1 GlnlleleclulysleulleGluLys 9
      ::::::::::::::::::::::::::::
Db 1054677 GAGATCTCGAAAAATTGATAGAAAG 1054651
```

RESULT 7

```
US-09-385-982-225
; Sequence 225, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 344
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 225
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(619)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-225
```

```
Alignment Scores:
Pred. No.: 14 Length: 619
Score: 35.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 85.37% Indels: 0
DB: 3 Gaps: 0
```

US-09-994-617-4 (1-9) x US-09-385-982-225 (1-619)

```
QY 1 GlnlleleclulysleulleGluLys 9
      ::::::::::::::::::::::::::::
Db 501 GAAGTCTAGAGAAATTATAGAAAA 527
```

RESULT 8

```
US-09-620-312D-328
; Sequence 328, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
```

NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 328  
LENGTH: 4663  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (292)..(3600)  
US-09-620-312D-328

Alignment Scores:  
Pred. No.: 160 Length: 4663  
Score: 35.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 66.67% Mismatches: 0  
Query Match: 85.37% Indels: 0  
DB: 4 Gaps: 0  
US-09-994-617-4 (1-9) x US-09-620-312D-328 (1-4663)

QY 1 GlnlleleGluLysLeuileGluLys 9  
DB 14 GAAGTCTAGAGAAATAATAGAAAA 40

## RESULT 9

US-09-557-884-1  
Sequence 1, Application US/09557884  
Patent No. 6506581  
GENERAL INFORMATION:  
APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Alignment Scores:  
Pred. No.: 3.13e+05 Length: 1830121  
Score: 34.00 Matches: 6

Percent Similarity: 88.89% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 1  
Query Match: 82.93% Indels: 0  
DB: 4 Gaps: 0

US-09-994-617-4 (1-9) x US-09-557-884-1 (1-1830121)

QY 1 GlnlleleGluLysLeuileGluLys 9  
DB 937257 CAATTTGCGAAAAACAGTTGAAAAA 937283

## RESULT 10

US-09-643-990A-1

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of

the Haemophilus influenzae Rd Genome, Fragments

Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186PIC1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:

Pred. No.: 3.13e+05 Length: 1830121

Score: 34.00 Matches: 6

Percent Similarity: 88.89% Conservative: 2

Best Local Similarity: 66.67% Mismatches: 1

Query Match: 82.93% Indels: 0

DB: 4 Gaps: 0

US-09-994-617-4 (1-9) x US-09-643-990A-1 (1-1830121)

QY 1 GlnlleleGluLysLeuileGluLys 9

```
Db 937257 CAAATTCGAAAAACAGTTGAAAA 937283
; Sequence 61, Application US/08850049
; Patent No. 5965726
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: METHOD OF ELIMINATING
; TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,049
; FILING DATE: 02-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,478
; FILING DATE: 26-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02908
; FILING DATE: 29-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 435
; FILING DATE: 27-MAR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MORRY, MARY J.
; REGISTRATION NUMBER: 34,398
; REFERENCE/DOCKET NUMBER: 2026-4006US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)751-6849
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; US-08-850-049-61
Alignment Scores:
Pred. No.: 2.25 Length: 56
Score: 33.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 80.49% Indels: 0
DB: 2 Gaps: 0
US-09-994-617-4 (1-9) x US-08-850-049-61 (1-56)
QY 1 GlnlleleGlulysLeulleGlulys 9
Db 11 CAAATCATCAGCAGCTGATCAGAAG 37
RESULT 13
US-08-050-478-60
; Sequence 60, Application US/08050478
; Patent No. 5972596
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: METHOD OF ELIMINATING
; TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,049
; FILING DATE: 02-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,478
; FILING DATE: 26-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02908
; FILING DATE: 29-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MORRY, MARY J.
; REGISTRATION NUMBER: 34,398
; REFERENCE/DOCKET NUMBER: 2026-4006US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)751-6849
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; US-08-850-049-60
Alignment Scores:
Pred. No.: 2.25 Length: 56
Score: 33.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 80.49% Indels: 0
DB: 2 Gaps: 0
US-09-994-617-4 (1-9) x US-08-850-049-60 (1-56)
QY 1 GlnlleleGlulysLeulleGlulys 9
Db 11 CAAATAATAGCAGCTGATTAATAAAG 37
RESULT 12
US-08-850-049-61
```

APPLICANT:  
TITLE OF INVENTION: METHOD OF ELIMINATING  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/050,478  
FILING DATE: 26-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/02908  
FILING DATE: 29-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/858,747  
FILING DATE: 27-MAR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MORRY, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELEPHONE: (212)751-6849  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-050-478-60  
Alignment Scores:  
Pred. No.: 2.25 Length: 56  
Score: 33.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 80.49% Indels: 0  
DB: 2 Gaps: 0  
US-09-994-617-4 (1-9) x US-08-050-478-60 (1-56)  
QY 1 GlnllelleglulysleutleGluLys 9  
Db 11 CAAATAAGAGCAGTAAATAAAAG 37  
RESULT 14  
US-08-050-478-61  
Sequence 61, Application US/08050478  
Patent No. 5972596  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: METHOD OF ELIMINATING  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK

COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/050,478  
FILING DATE: 26-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/02908  
FILING DATE: 29-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/858,747  
FILING DATE: 27-MAR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MORRY, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELEPHONE: (212)751-6849  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-050-478-61  
Alignment Scores:  
Pred. No.: 2.25 Length: 56  
Score: 33.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 80.49% Indels: 0  
DB: 2 Gaps: 0  
US-09-994-617-4 (1-9) x US-08-050-478-61 (1-56)  
QY 1 GlnllelleglulysleutleGluLys 9  
Db 11 CAAATCATGAGCAGCTGATCAAGAAG 37  
RESULT 15  
US-09-414-117-60  
Sequence 60, Application US/09414117  
Patent No. 6291664  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: METHOD OF ELIMINATING  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/414,117

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/850,049
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCV/US93/02908
; FILING DATE: 29-MAR-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/858,747
; FILING DATE: 27-MAR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MORRY, MARY J.
; REGISTRATION NUMBER: 34,398
; REFERENCE/DOCKET NUMBER: 2026-4006US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
;
US-09-414-117-60

Alignment Scores:
Pred. No.: 2.25 Length: 56
Score: 33.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 80.49% Indels: 0
DB: 3 Gaps: 0

US-09-994-617-4 (1-9) x US-09-414-117-60 (1-56)
Qy 1 GlnIleIleGluLysLeuIleGluLys 9
Db 11 CAATATATAGAGCAGTTTATATAAAAG 37

```

Search completed: August 6, 2003, 13:10:53  
Job time : 889 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 6, 2003, 13:12:51 ; Search time 202 Seconds  
(without alignments)  
91.916 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIEKLEK 9

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPRO\_spool/US09994617/runat\_29072003\_160425\_7748/app\_query.fasta\_1.199  
-DB=PublishedApplications\_NA -Qfmt=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09994617@cgn1.1.333 @runat\_29072003\_160425\_7748  
-NCPU=6 -ICPU=3 -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	92.7	1088	13	US-10-027-632-117623 Sequence 117623,

2	37	90.2	8509	9	US-09-826-205-1	Sequence 1, Appli
3	35	85.4	517	14	US-10-106-698-1393	Sequence 1393, Ap
4	35	85.4	539	14	US-10-102-524-458	Sequence 458, App
5	35	85.4	619	11	US-09-871-161-225	Sequence 225, App
6	35	85.4	788	13	US-10-027-632-139223	Sequence 139223,
7	35	85.4	788	13	US-10-027-632-139224	Sequence 139224,
8	35	85.4	1051	9	US-09-925-301-41	Sequence 41, Appl
9	35	85.4	4663	14	US-10-037-370-328	Sequence 328, App
10	35	85.4	640681	10	US-09-790-988-1	Sequence 1, Appli
11	34	82.9	274	10	US-09-878-574-6773	Sequence 6773, Ap
12	34	82.9	1082	13	US-10-027-632-250777	Sequence 250777,
13	34	82.9	1082	13	US-10-027-632-250778	Sequence 250778,
14	34	82.9	1614	10	US-09-344-882-21	Sequence 21, Appl
15	34	82.9	1614	14	US-10-293-865-21	Sequence 21, Appl
16	34	82.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
17	33	80.5	56	10	US-09-943-722-60	Sequence 60, Appl
18	33	80.5	56	10	US-09-943-722-61	Sequence 61, Appl
19	33	80.5	337	11	US-09-918-995-29836	Sequence 29836, A
20	33	80.5	431	9	US-09-604-287A-115	Sequence 115, App
21	33	80.5	431	10	US-09-339-338-115	Sequence 115, App
22	33	80.5	431	11	US-09-551-621-115	Sequence 115, App
23	33	80.5	431	13	US-10-007-805-115	Sequence 115, App
24	33	80.5	431	14	US-10-076-822-115	Sequence 115, App
25	33	80.5	483	11	US-09-918-995-2878	Sequence 2878, Ap
26	33	80.5	585	13	US-10-027-632-219893	Sequence 219893,
27	33	80.5	612	13	US-10-027-632-114144	Sequence 114144,
28	33	80.5	2000	10	US-09-938-842A-3091	Sequence 3091, Ap
29	33	80.5	2348	13	US-10-097-997-1	Sequence 1, Appli
30	33	80.5	2467	9	US-09-872-733-3	Sequence 3, Appli
31	33	80.5	2467	14	US-10-263-020-3	Sequence 3, Appli
32	33	80.5	2507	9	US-09-872-733-2	Sequence 2, Appli
33	33	80.5	2507	14	US-10-263-020-2	Sequence 2, Appli
34	33	80.5	2577	11	US-09-952-060-1	Sequence 1, Appli
35	33	80.5	2577	11	US-09-952-060-3	Sequence 3, Appli
36	33	80.5	2601	9	US-09-735-487-7	Sequence 7, Appli
37	33	80.5	2601	9	US-09-735-487-9	Sequence 9, Appli
38	33	80.5	2601	9	US-09-735-487-11	Sequence 11, Appl
39	33	80.5	2601	9	US-09-735-487-13	Sequence 13, Appl
40	33	80.5	2650	11	US-09-952-060-5	Sequence 5, Appli
41	33	80.5	2650	11	US-09-952-060-7	Sequence 7, Appli
42	33	80.5	4053	11	US-09-952-060-34	Sequence 34, Appl
43	33	80.5	4307	10	US-09-999-183-1	Sequence 1, Appli
44	33	80.5	4307	10	US-09-999-183-2	Sequence 2, Appli
45	33	80.5	4338	9	US-09-872-733-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-10-027-632-117623  
; Sequence 117623, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720

US-10-027-632-117623

Alignment Scores:				
Pred. No.:	71.2	Length:	517	
Score:	35.00	Matches:	6	
Percent Similarity:	100.00%	Conservative:	3	
Best Local Similarity:	66.67%	Mismatches:	0	
Query Match:	85.37%	Indels:	0	
DB:	14	Gaps:	0	

US-09-994-617-4 (1-9) x US-10-106-698-1393 (1-517)

**Oy**            1 GlnIleIeGluLysLeuIfleGluLys 9  
               : : : : :  
**Dd**            455 GAAGTTC TAGAGAAATTATAGAAAAA 429

```

RESULT 4
US-10-102-524-458/c
; Sequence 458, Application US/10402524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT OF KIDNEY CANCER
; FILE REFERENCE: 210421-572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 458
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 497
; OTHER INFORMATION: n = A,T,C or G
US-10-102-524-458

```

Alignment Scores:					
Pred. No.:	74.9	Length:	539		
Score:	35.00	Matches:	6		
Percent Similarity:	100.00%	Conservative:	3		
Best Local Similarity:	66.67%	Mismatches:	0		
Query Match:	85.37%	Indels:	0		
DB:	14	Gaps:	0		

US-09-994-617-4 (1-9) x US-10-102-524-458 (1-539)

**Qy** 1 GlnIleIleGluLysLeuIleGluLys 9  
:::~::~|||||||~|||||  
**Db** 375 GAAGTTCTAGAGAAATTAATAGAAAA 349

```

RESULT 5
; Sequence 225, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O. A. ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS; II
; FILE REFERENCE: CCDNA-260XX;

```







```

RESULT 12
US-10-027-632-250777/c
; Sequence 250777, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250777
; LENGTH: 1082
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250777

Alignment Scores:
Pred. No.: 296 Length: 1082
Score: 34.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 13 Gaps: 0

US-09-994-617-4 (1-9) x US-10-027-632-250777 (1-1082)
QY 2 llellegluLysLeuilegluLys 9
Db 399 CTTATAGAAAAAATTATTGAAAAA 376

RESULT 13
US-10-027-632-250778/c
; Sequence 250778, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250778

```

```

; LENGTH: 1082
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250778

Alignment Scores:
Pred. No.: 296 Length: 1082
Score: 34.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 13 Gaps: 0

US-09-994-617-4 (1-9) x US-10-027-632-250778 (1-1082)
QY 2 llellegluLysLeuilegluLys 9
Db 399 CTTATAGAAAAAATTATTGAAAAA 376

RESULT 14
US-09-344-882-21
; Sequence 21, Application US/09344882
; Patent No. US20020162137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 21
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-21

Alignment Scores:
Pred. No.: 486 Length: 1614
Score: 34.00 Matches: 5
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 55.56% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 10 Gaps: 0

US-09-994-617-4 (1-9) x US-09-344-882-21 (1-1614)
QY 1 GinllellegluLysLeuilegluLys 9
Db 981 AAGTTGTGGAGAAGTTAGTAGAGAAG 1007

RESULT 15
US-10-293-865-21
; Sequence 21, Application US/10293865
; Publication NO. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert

```

; APPLICANT: Schnable, Patrick S  
; APPLICANT: Ke, Jinshan  
; APPLICANT: Johnson, Jerry L  
; APPLICANT: Allred, Carolyn C  
; APPLICANT: Fatland, Beth  
; APPLICANT: Lutziger, Isabelle  
; APPLICANT: Wen, Tsui-Jung  
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and  
; FILE OF INVENTION: Acetyl CoA Levels in Plants  
; FILE REFERENCE: 217113  
; CURRENT APPLICATION NUMBER: US/10/293,865  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/344,882  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 60/090,717  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 21  
; LENGTH: 1614  
; TYPE: DNA  
; ORGANISM: Arabidopsis Thaliana  
US-10-293-865-21

Alignment Scores:  
Pred. No.: 486 Length: 1614  
Score: 34.00 Matches: 5  
Percent Similarity: 100.00% Conservative: 4  
Best Local Similarity: 55.56% Mismatches: 0  
Query Match: 82.93% Indels: 0  
DB: 14 Gaps: 0

US-09-994-617-4 (1-9) x US-10-293-865-21 (1-1614)

Qy 1 GlnlleleGlutylsleulleGlutys 9  
Db 981 AAGGTGTGGAGGAGTTAGTAGAGAG 1007

Search completed: August 6, 2003, 14:03:05  
Job time : 257 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 6, 2003, 11:44:09 ; Search time 3450 Seconds  
(without alignments)  
85.481 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIIIEKLIK 9

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 32997241 seqs, 16383922548 residues

Total number of hits satisfying chosen parameters: 65994482

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-DB=Pending\_Patents\_NA\_Main -QFMT=fastap -SUFFIX=rnnp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=diom62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09994617 -CGN\_1\_1\_5832 -runat.29072003.160424.7696 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending\_Patents\_NA\_Main: \*  
1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.old.\*  
3: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*

29: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US098D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US099E\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US099F\_COMB.seq.\*  
44: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US103A\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US103B\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US104A\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US104B\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6023A\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6023B\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*  
86: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*  
87: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq.\*  
88: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq.\*  
89: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq.\*  
90: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq.\*  
91: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq.\*  
92: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq.\*  
93: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq.\*  
94: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq.\*  
95: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq.\*  
96: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq.\*  
97: /cgn2\_6/ptodata/1/pna/US6042\_COMB.seq.\*  
98: /cgn2\_6/ptodata/1/pna/US6043\_COMB.seq.\*  
99: /cgn2\_6/ptodata/1/pna/US6044\_COMB.seq.\*  
100: /cgn2\_6/ptodata/1/pna/US6045\_COMB.seq.\*  
101: /cgn2\_6/ptodata/1/pna/US6046\_COMB.seq.\*



```
; FILE REFERENCE: 12804-006001
; CURRENT APPLICATION NUMBER: US/10/336,566
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US 10/093,953
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 10447
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated vector sequence-pGAL/IC25
US-10-336-566-10

Alignment Scores:
Pred. No.:      2.41e+03      Length:      10447
Score:          41.00         Matches:      9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%       Indels:       0
DB:             50           Gaps:         0

US-09-994-617-4 (1-9) x US-10-336-566-10 (1-10447)
QY      1 GlnlIlelGluLysLeuileGluLys 9
      |||||
Db      3415 CAAATAAGAGAACTAATAAGAAAG 3441

RESULT 4
US-10-336-566-11
; Sequence 11, Application US/10336566
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE REFERENCE: 12804-006001
; CURRENT APPLICATION NUMBER: US/10/336,566
; PRIOR APPLICATION NUMBER: US 10/093,953
; PRIOR FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 10447
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated vector sequence-pGAL/IC2
US-10-336-566-11
```

```
Alignment Scores:
Pred. No.:      2.41e+03      Length:      10447
Score:          41.00         Matches:      9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%       Indels:       0
DB:             50           Gaps:         0

US-09-994-617-4 (1-9) x US-10-336-566-11 (1-10447)
QY      1 GlnlIlelGluLysLeuileGluLys 9
      |||||
Db      3415 CAAATAAGAGAACTAATAAGAAAG 3441

RESULT 5
US-10-336-566-12
; Sequence 12, Application US/10336566
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE REFERENCE: 12804-006001
; CURRENT APPLICATION NUMBER: US/10/336,566
; PRIOR APPLICATION NUMBER: US 10/093,953
; PRIOR FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 12
; LENGTH: 10447
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated vector sequence-pGAL/IC48
US-10-336-566-12

Alignment Scores:
Pred. No.:      2.41e+03      Length:      10447
Score:          41.00         Matches:      9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%       Indels:       0
DB:             50           Gaps:         0

US-09-994-617-4 (1-9) x US-10-336-566-12 (1-10447)
QY      1 GlnlIlelGluLysLeuileGluLys 9
      |||||
Db      3415 CAAATAAGAGAACTAATAAGAAAG 3441

RESULT 6
US-10-336-566-13
; Sequence 13, Application US/10336566
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE REFERENCE: 12804-006001
; CURRENT APPLICATION NUMBER: US/10/336,566
```

```
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US 10/093,953
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/798,675
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 10447
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated vector sequence-pGAL/IC90
US-10-336-566-13

Alignment Scores:
Pred. No.:      2,41e+03      Length:      10447
Score:          41.00         Matches:      9
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:     100.00%      Indels:        0
DB:              50          Gaps:          0

US-09-994-617-4 (1-9) x US-10-336-566-13 (1-10447)

QY      1 Glnllelleglulysleulleglulys 9
      |||:::|||||:::|||||
Db      3415 CAAATAATAGAGAACTAATAGAAAG 3441

RESULT 7
US-09-634-306B-117623
; Sequence 117623, Application US/09634306B
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/09/634,306B
; CURRENT FILING DATE: 2002-02-21
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117623
; LENGTH: 1088
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1088)
; OTHER INFORMATION: n = A,T,C or G
US-09-634-306B-117623

Alignment Scores:
Pred. No.:      892          Length:      1088
Score:          38.00         Matches:      7
Percent Similarity: 100.00%    Conservative: 2
Best Local Similarity: 77.78%  Mismatches:      0
Query Match:     92.68%      Indels:        0
DB:              44          Gaps:          0

US-09-994-617-4 (1-9) x US-10-027-632-117623 (1-1088)

QY      1 Glnllelleglulysleulleglulys 9
      |||:::|||||:::|||||
Db      421 CAGGTGATAGAGAAATCATTTGAAAAA 447

RESULT 9
US-09-947-911-294/C
; Sequence 294, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED ON
; FILE REFERENCE: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
US-09-634-306B-117623
```

```
Alignment Scores:
Pred. No.:      892          Length:      1088
Score:          38.00         Matches:      7
Percent Similarity: 100.00%    Conservative: 2
Best Local Similarity: 77.78%  Mismatches:      0
Query Match:     92.68%      Indels:        0
DB:              27          Gaps:          0

US-09-994-617-4 (1-9) x US-09-634-306B-117623 (1-1088)

QY      1 Glnllelleglulysleulleglulys 9
      |||:::|||||:::|||||
Db      421 CAGGTGATAGAGAAATCATTTGAAAAA 447

RESULT 8
US-10-027-632-117623
; Sequence 117623, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117623
; LENGTH: 1088
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1088)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-117623

Alignment Scores:
Pred. No.:      892          Length:      1088
Score:          38.00         Matches:      7
Percent Similarity: 100.00%    Conservative: 2
Best Local Similarity: 77.78%  Mismatches:      0
Query Match:     92.68%      Indels:        0
DB:              44          Gaps:          0

US-09-994-617-4 (1-9) x US-10-027-632-117623 (1-1088)

QY      1 Glnllelleglulysleulleglulys 9
      |||:::|||||:::|||||
Db      421 CAGGTGATAGAGAAATCATTTGAAAAA 447

RESULT 9
US-09-947-911-294/C
; Sequence 294, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPS) LOCATED ON
; FILE REFERENCE: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001300
; CURRENT APPLICATION NUMBER: US/09/947,911
US-09-947-911-294/C
```



; CURRENT FILING DATE: 2001-09-07  
; NUMBER OF SEQ ID NOS: 368  
; SEQ ID NO 294  
; LENGTH: 2347870  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(2347870)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-947-911-294

Alignment Scores:  
Pred. No.: 3.82e+06 2347870  
Score: 38.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 92.68% Indels: 0  
DB: 40 Gaps: 0

US-09-994-617-4 (1-9) x US-09-947-911-294 (1-2347870)

QY 1 GlnllelleGluLysLeulleclulys 9  
|||||  
DB 1944106 CAAATAATTGAGAAATTAATTCGACAAA 1944080

RESULT 10

US-09-404-520-9568/c

; Sequence 9568, Application US/09404520

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Ghodssi, Azita

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: McIninch, James

; APPLICANT: Timberlake, William E.

; APPLICANT: Yu, Jaehyuk

; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses Thereof

; FILE REFERENCE: 38-10(15498)A

; CURRENT APPLICATION NUMBER: US/09/404,520

; CURRENT FILING DATE: 1999-09-23

; NUMBER OF SEQ ID NOS: 44345

; SEQ ID NO 9568

; LENGTH: 374

; TYPE: DNA

; ORGANISM: Aspergillus nidulans

US-09-404-520-9568

Alignment Scores:

Pred. No.: 455 Length: 374  
Score: 37.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 90.24% Indels: 0  
DB: 21 Gaps: 0

US-09-994-617-4 (1-9) x US-09-404-520-9568 (1-374)

QY 1 GlnllelleGluLysLeulleclulys 9  
|||||  
DB 186 CAAATTATCGAAGATTCGAGAAG 160

RESULT 11

US-60-138-103-22569/c

; Sequence 22569, Application US/60138103

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Timberlake, William E.

; TITLE OF INVENTION: ASPERGILLUS NIDULANS GENOME SEQUENCE AND USES THEREOF

; FILE REFERENCE: 38-10(15485)D

; CURRENT APPLICATION NUMBER: US/60/138,103

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 28006

; SEQ ID NO 22569

; LENGTH: 374  
; TYPE: DNA  
; ORGANISM: Aspergillus nidulans  
US-60-138-103-22569

Alignment Scores:

Pred. No.: 455 Length: 374  
Score: 37.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 90.24% Indels: 0  
DB: 67 Gaps: 0

US-09-994-617-4 (1-9) x US-60-138-103-22569 (1-374)

QY 1 GlnllelleGluLysLeulleclulys 9  
|||||  
DB 186 CAAATTATCGAAGATTCGAGAAG 160

RESULT 12

US-10-431-652-2627

; Sequence 2627, Application US/10431652

; GENERAL INFORMATION:

; APPLICANT: Breton, Gary L.

; APPLICANT: Bush, David

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: PATH03-08

; CURRENT APPLICATION NUMBER: US/10/431,652

; CURRENT FILING DATE: 2003-05-06

; PRIOR APPLICATION NUMBER: US 09/328,352

; PRIOR FILING DATE: 1999-06-04

; PRIOR APPLICATION NUMBER: US 60/088,701

; PRIOR FILING DATE: 1998-06-09

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 2627

; LENGTH: 1482

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-10-431-652-2627

Alignment Scores:

Pred. No.: 2.05e+03 Length: 1482  
Score: 37.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 90.24% Indels: 0  
DB: 53 Gaps: 0

US-09-994-617-4 (1-9) x US-10-431-652-2627 (1-1482)

QY 1 GlnllelleGluLysLeulleclulys 9  
||| |||||  
DB 928 CAAATTATGAAAAAAGTTCGAAAAA 954

RESULT 13

US-09-826-206-2

; Sequence 2, Application US/09826206

; GENERAL INFORMATION:

; APPLICANT: D'Elia, John

; TITLE OF INVENTION: Ketoqulonigenium Shuttle Vectors

; FILE REFERENCE: 1533.1100001

; CURRENT APPLICATION NUMBER: US/09/826,206

; CURRENT FILING DATE: 2001-04-05

; PRIOR APPLICATION NUMBER: US 60/194,625

; PRIOR FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 8509

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; NAME/KEY: misc\_feature  
; OTHER INFORMATION: padm291  
US-09-826-206-2

Alignment Scores: Length: 8509  
Pred. No.: 1.38e+04  
Score: 37.00  
Percent Similarity: 100.00%  
Best Local Similarity: 77.78%  
Query Match: 90.24%  
DB: 34

US-09-994-617-4 (1-9) x US-09-826-206-2 (1-8509)

Qy 1 GlnIleIleGluLysLeuIleGluLys 9  
|||||:|||||:|||||:|||||  
Db 2650 CAATCTTGAAAAAATTGCTTGAAG 2676

## RESULT 14

US-09-826-206A-2  
; Sequence 2, Application US/09826206A  
; GENERAL INFORMATION:  
; APPLICANT: D'Elia, John  
; TITLE OF INVENTION: Ketogulonigenium Shuttle Vectors  
; FILE REFERENCE: 1533.1100001  
; CURRENT APPLICATION NUMBER: US/09/826,206A  
; CURRENT FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,625  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 8509  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: padm291  
US-09-826-206A-2

Alignment Scores: Length: 8509  
Pred. No.: 1.38e+04  
Score: 37.00  
Percent Similarity: 100.00%  
Best Local Similarity: 77.78%  
Query Match: 90.24%  
DB: 34

US-09-994-617-4 (1-9) x US-09-826-206A-2 (1-8509)

Qy 1 GlnIleIleGluLysLeuIleGluLys 9  
|||||:|||||:|||||:|||||  
Db 2650 CAATCTTGAAAAAATTGCTTGAAG 2676

## RESULT 15

US-60-230-445-638/c  
; Sequence 638, Application US/60230445  
; GENERAL INFORMATION:  
; APPLICANT: Beasley, Ellen  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL000765  
; CURRENT APPLICATION NUMBER: US/60/230,445  
; CURRENT FILING DATE: 2000-09-06  
; NUMBER OF SEQ ID NOS: 3051  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 638  
; LENGTH: 32338  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(32338)

; OTHER INFORMATION: n = A,T,C or G  
US-60-230-445-638

Alignment Scores: Length: 32338  
Pred. No.: 5.96e+04  
Score: 37.00  
Percent Similarity: 100.00%  
Best Local Similarity: 77.78%  
Query Match: 90.24%  
DB: 77

US-09-994-617-4 (1-9) x US-60-230-445-638 (1-32338)

Qy 1 GlnIleIleGluLysLeuIleGluLys 9  
|||||:|||||:|||||:|||||  
Db 17996 CAGATTATAGAAAAATTAGTGGAAAGA 17970

Search completed: August 6, 2003, 13:59:10  
Job time : 3622 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 6, 2003, 12:15:46 ; Search time 80 Seconds  
(without alignments)  
69.673 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIEKLEIK 9

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 502295 seqs, 309655796 residues

Total number of hits satisfying chosen parameters: 1004590

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09994617/runat\_29072003.160424.7726/app\_query.fasta.1.199  
-DB=Pending\_Patents\_NA\_New -QFMT=fastap -SUFFIX=rnnp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09994617@cgn.1.1.43@runat\_29072003.160424.7726 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT\_DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	85.4	4663	US-10-105-837-328	Sequence 328, App
2	35	85.4	4663	US-10-286-897-650	Sequence 650, App
C 3	34	82.9	4040	US-60-487-610-445	Sequence 445, App
C 4	34	82.9	124326	US-60-487-610-19458	Sequence 19458, A
5	34	82.9	125896	US-60-485-450-11939	Sequence 11939, A
C 6	34	82.9	135024	US-60-487-610-19904	Sequence 19904, A
C 7	34	82.9	13831263	US-09-947-914-41	Sequence 41, Appl
8	32	78.0	639	US-10-603-113-8381	Sequence 8381, Ap
9	32	78.0	903	US-10-603-113-1743	Sequence 1743, Ap
10	32	78.0	1149	US-10-603-114-67	Sequence 67, Appl
11	32	78.0	2370	US-10-603-113-5191	Sequence 5191, Ap

12	78.0	2401	6	US-10-302-172-523	Sequence 523, App
13	78.0	3243	6	US-10-603-114-1875	Sequence 1875, Ap
14	78.0	7971	7	US-60-478-196-15	Sequence 15, Appl
15	78.0	41554	7	US-60-485-450-12179	Sequence 12179, A
16	78.0	42314	7	US-60-487-610-13758	Sequence 13758, A
17	78.0	73378	7	US-60-487-610-19464	Sequence 19464, A
18	78.0	1748349	5	US-09-947-914-48	Sequence 48, Appl
19	78.0	1791310	5	US-09-947-914-46	Sequence 46, Appl
C 20	78.0	1791310	5	US-09-947-914-46	Sequence 46, Appl
C 21	78.0	4813087	5	US-09-947-914-75	Sequence 75, Appl
C 22	78.0	8059021	5	US-09-947-914-53	Sequence 53, Appl
23	78.0	13831263	5	US-09-947-914-41	Sequence 41, Appl
C 24	78.0	201	7	US-60-485-450-18190	Sequence 18190, A
C 25	78.0	636	6	US-10-273-573-2442	Sequence 2442, Ap
26	78.0	720	6	US-10-357-930-15177	Sequence 15177, A
C 27	78.0	1068	6	US-10-273-573-4600	Sequence 4600, Ap
C 28	78.0	1233	6	US-10-603-113-4353	Sequence 4353, Ap
C 29	78.0	1311	7	US-60-479-073-442	Sequence 442, App
C 30	78.0	1422	7	US-60-479-073-442	Sequence 442, App
C 31	78.0	1553	6	US-10-291-172-96	Sequence 96, Appl
C 32	78.0	2265	7	US-60-478-196-1190	Sequence 1190, Ap
C 33	78.0	2265	7	US-60-478-196-2190	Sequence 2190, Ap
34	78.0	2278	6	US-10-291-172-472	Sequence 472, App
35	78.0	3014	5	US-09-784-553C-1	Sequence 1, Appl
36	78.0	3014	6	US-10-209-201C-1	Sequence 1, Appl
C 37	78.0	3156	6	US-10-603-114-3971	Sequence 3971, Ap
C 38	78.0	3588	5	US-09-915-706B-1	Sequence 1, Appl
39	78.0	4195	7	US-60-485-450-179	Sequence 179, App
40	78.0	5065	6	US-10-357-930-24299	Sequence 24299, A
41	78.0	5761	6	US-10-302-172-23	Sequence 23, Appl
C 42	78.0	8265	7	US-60-478-196-190	Sequence 190, App
C 43	78.0	9715	6	US-10-279-992A-6	Sequence 6, Appl
44	78.0	9876	6	US-10-279-992A-3	Sequence 3, Appl
45	78.0	9984	6	US-10-279-992A-5	Sequence 5, Appl

#### ALIGNMENTS

RESULT 1  
US-10-105-837-328  
; Sequence 328, Application US/10105837  
; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Drmanac, Radojko T.  
; TITLE OF INVENTION: Novel Nucleic Acids and  
; FILE REFERENCE: 784CIP2BDIVA  
; CURRENT APPLICATION NUMBER: US/10/105,837  
; CURRENT FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 328  
; LENGTH: 4663  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (292)..(3600)  
US-10-105-837-328

Alignment Scores: 114 Length: 4663  
Pred. No.: 35.00 Matches: 6  
Score: 100.00%  
Percent Similarity: 100.00% Conservative: 3

```
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 85.37% Indels: 0
DB: 6 Gaps: 0

US-09-994-617-4 (1-9) x US-10-105-837-328 (1-4663)

Qy 1 GlnlleleGluLysLeulleGluLys 9
Db 14 GAAGTTCTAGAGAAATTAATAGAAAAA 40

RESULT 2
US-10-286-897-650
; Sequence 650, Application US/10286897
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acid and Polypeptides
; FILE REFERENCE: 784FLPCT
; CURRENT APPLICATION NUMBER: US/10/286,897
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US/09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US/09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US/09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US/09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US/09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US/09/662,191
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: US/09/693,036
; PRIOR FILING DATE: 2000-10-19
; PRIOR APPLICATION NUMBER: US/09/727,344
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 7143
; SOFTWARE: pt_FL_genes_b Versions 1.0
; SEQ ID NO 650
; LENGTH: 4663
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (292)..(3600)
US-10-286-897-650

Alignment Scores:
Pred. No.: 114 Length: 4663
Score: 35.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 85.37% Indels: 0
DB: 6 Gaps: 0

US-09-994-617-4 (1-9) x US-10-286-897-650 (1-4663)

Qy 1 GlnlleleGluLysLeulleGluLys 9
Db 14 GAAGTTCTAGAGAAATTAATAGAAAAA 40

RESULT 3
US-60-487-610-445/C
; Sequence 445, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 445
; LENGTH: 4040
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-445

Alignment Scores:
Pred. No.: 165 Length: 4040
Score: 34.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 7 Gaps: 0

US-09-994-617-4 (1-9) x US-60-487-610-445 (1-4040)

Qy 1 GlnlleleGluLysLeulleGluLys 9
Db 2270 AAGTTATAGAAAACTCTTAGAAAAA 2244

RESULT 4
US-60-487-610-19458/C
; Sequence 19458, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19458
; LENGTH: 124326
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-19458

Alignment Scores:
Pred. No.: 6,37e+03 Length: 124326
Score: 34.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 7 Gaps: 0

US-09-994-617-4 (1-9) x US-60-487-610-19458 (1-124326)

Qy 1 GlnlleleGluLysLeulleGluLys 9
Db 67337 AAGTTATAGAAAACTCTTAGAAAAA 67311

RESULT 5
US-60-485-450-11939
; Sequence 11939, Application US/60485450
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: CHANG, Sheng-yung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/485,450
; CURRENT FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11939
; LENGTH: 125896
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(125896)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-09-994-617-4 (1-9) x US-09-994-617-4 (1-125896)

Alignment Scores:
Pred. No.: 6.45e+03 Length: 125896
Score: 34.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 7 Gaps: 0

US-09-994-617-4 (1-9) x US-60-485-450-11939 (1-125896)

QY 1 GlnlleGluLysLeuileGluLys 9
DB 111802 CAGATAATTAATAAACTGATTGAAGA 111828

RESULT 6
US-60-487-610-19904/c
; Sequence 19904, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; NUMBER OF SEQ ID NOS: 97101
; CURRENT FILING DATE: 2003-07-17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19904
; LENGTH: 135024
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-19904

Alignment Scores:
Pred. No.: 6.95e+03 Length: 135024
Score: 34.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 7 Gaps: 0

US-09-994-617-4 (1-9) x US-60-487-610-19904 (1-135024)

QY 1 GlnlleGluLysLeuileGluLys 9
DB 40831 CAATTTGTTGAGAAATTAATTAAGAA 40805

RESULT 7
US-09-947-914-41/c
; Sequence 41, Application US/09947914
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
; TITLE OF INVENTION: 21, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL001298
; CURRENT APPLICATION NUMBER: US/09/947,914
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 75
; SEQ ID NO 41
; LENGTH: 13831263
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(13831263)
```

```
; OTHER INFORMATION: n = A,T,C or G
US-09-947-914-41

Alignment Scores:
Pred. No.: 6.21e+05 Length: 13831263
Score: 34.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 82.93% Indels: 0
DB: 5 Gaps: 0

US-09-994-617-4 (1-9) x US-09-947-914-41 (1-13831263)

QY 2 llelleGluLysLeuileGluLys 9
DB 8550687 ATAATAGAGAAATCATAGAAAAA 8550664

RESULT 8
US-10-603-113-8381
; Sequence 8381, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603,113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248,796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 8381
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-8381

Alignment Scores:
Pred. No.: 65.3 Length: 639
Score: 32.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 78.05% Indels: 0
DB: 6 Gaps: 0

US-09-994-617-4 (1-9) x US-10-603-113-8381 (1-639)

QY 1 GlnlleGluLysLeuileGluLys 9
DB 130 AAACATCATACAAAAAACTAATTGAAAAG 156

RESULT 9
US-10-603-113-1743
; Sequence 1743, Application US/10603113
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/10/603,113
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US/09/248,796
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 1743
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-603-113-1743

Alignment Scores:
Pred. No.: 94.5 Length: 903
Score: 32.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
```

Best Local Similarity: 66.67% Mismatches: 0  
Query Match: 78.05% Indels: 0  
DB: 6 Gaps: 0

US-09-994-617-4 (1-9) x US-10-603-113-1743 (1-903)

QY 1 GlnIleleGlulysLeuileGluLys 9

DB 532 AAAATTATTCACAAGATTATAGAAAG 558

## RESULT 10

US-10-603-114-67

; Sequence 67, Application US/10603114

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/10/603,114

; PRIOR FILING DATE: 2003-06-24

; PRIOR APPLICATION NUMBER: US/09/543,681

; PRIOR FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 67

; LENGTH: 1149

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-10-603-114-67

## Alignment Scores:

Pred. No.:	122	Length:	1149
Score:	32.00	Matches:	7
Percent Similarity:	87.50%	Conservative:	0
Best Local Similarity:	87.50%	Mismatches:	1
Query Match:	78.05%	Indels:	0
DB:	6	Gaps:	0

US-09-994-617-4 (1-9) x US-10-603-114-67 (1-1149)

QY 2 IleileleGlulysLeuileGluLys 9

DB 124 ATTTTGAAGAACTAATTGAAAG 147

## RESULT 11

US-10-603-113-5191

; Sequence 5191, Application US/10603113

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/10/603,113

; PRIOR FILING DATE: 2003-06-24

; PRIOR APPLICATION NUMBER: US/09/248,796

; PRIOR FILING DATE: 1999-02-12

; NUMBER OF SEQ ID NOS: 28206

; SEQ ID NO 5191

; LENGTH: 2370

; TYPE: DNA

; ORGANISM: Candida albicans

US-10-603-113-5191

## Alignment Scores:

Pred. No.:	264	Length:	2370
Score:	32.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	4
Best Local Similarity:	55.56%	Mismatches:	0
Query Match:	78.05%	Indels:	0
DB:	6	Gaps:	0

US-09-994-617-4 (1-9) x US-10-603-113-5191 (1-2370)

QY 1 GlnIleleGlulysLeuileGluLys 9

DB 730 CAATTATTAAGAATAATGTCGAGAA 756

## RESULT 12

US-10-302-172-523

; Sequence 523, Application US/10302172

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: Novel Arginine-rich Protein-like Nucleic Acids and

; FILE REFERENCE: 803.1CNCp

; CURRENT APPLICATION NUMBER: US/10/302,172

; PRIOR FILING DATE: 2002-11-21

; PRIOR APPLICATION NUMBER: US 10/225,251

; PRIOR FILING DATE: 2002-08-20

; PRIOR APPLICATION NUMBER: PCT US02/05095

; PRIOR FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: US 09/799,451

; NUMBER OF SEQ ID NOS: 950

; SOFTWARE: pt\_FL\_genes Version 2.0

; SEQ ID NO 523

; LENGTH: 2401

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (170)..(2116)

US-10-302-172-523

## Alignment Scores:

Pred. No.:	268	Length:	2401
Score:	32.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	78.05%	Indels:	0
DB:	6	Gaps:	0

US-09-994-617-4 (1-9) x US-10-302-172-523 (1-2401)

QY 1 GlnIleleGlulysLeuileGlu 8

DB 1709 CAGATCATTAAGAATAATGAG 1732

## RESULT 13

US-10-603-114-1875

; Sequence 1875, Application US/10603114

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/10/603,114

; PRIOR FILING DATE: 2003-06-24

; PRIOR APPLICATION NUMBER: US/09/543,681

; PRIOR FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 1875

; LENGTH: 3243

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-10-603-114-1875

## Alignment Scores:

Pred. No.:	369	Length:	3243
Score:	32.00	Matches:	5
Percent Similarity:	88.89%	Conservative:	3

Best Local Similarity: 55.56%  
Query Match: 78.05%  
DB: Mismatches: 1  
Indels: 0  
Gaps: 0

US-09-994-617-4 (1-9) x US-10-603-114-1875 (1-3243)

Qy 1 GlnlleleGluLysLeuilleGluLys 9

Db 229 CAAAGTGTGAGAAAGATTATTGAGAAA 255

RESULT 14

US-60-478-196-15

; Sequence 15, Application US/60478196

; GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Lemieux, Sebastien

; APPLICANT: Hu, Wenqi

; APPLICANT: Rogner, Terry

; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF ASPERGILLUS FUMIGATUS AND ME

; FILE REFERENCE: 10182-026-888

; CURRENT APPLICATION NUMBER: US/60/478,196

; CURRENT FILING DATE: 2003-06-13

; NUMBER OF SEQ ID NOS: 4000

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 15

; LENGTH: 7971

; TYPE: DNA

; ORGANISM: Aspergillus fumigatus

US-60-478-196-15

Alignment Scores:

Pred. No.: 964 Length: 7971

Score: 32.00 Matches: 6

Percent Similarity: 100.00% Conservative: 2

Best Local Similarity: 75.00% Mismatches: 0

Query Match: 78.05% Indels: 0

DB: 7 Gaps: 0

US-09-994-617-4 (1-9) x US-60-478-196-15 (1-7971)

Qy 1 GlnlleleGluLysLeuilleGlu 8

Db 6504 CAAATGGTGGAAAAGTTGATCGAG 6527

RESULT 15

US-60-485-450-12179

; Sequence 12179, Application US/60485450

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele

; APPLICANT: CHANG, Sheng-Yung

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C

; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CLO01470

; CURRENT APPLICATION NUMBER: US/60/485,450

; CURRENT FILING DATE: 2003-07-09

; NUMBER OF SEQ ID NOS: 47859

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12179

; LENGTH: 41554

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(41554)

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-

US-60-485-450-12179

Alignment Scores:

Pred. No.: 5.61e+03 Length: 41554

Score: 32.00 Matches: 6

Percent Similarity: 88.89%  
Best Local Similarity: 66.67%  
Query Match: 78.05%  
DB: 7  
Mismatches: 1  
Indels: 0  
Gaps: 0

US-09-994-617-4 (1-9) x US-60-485-450-12179 (1-41554)

Qy 1 GlnlleleGluLysLeuilleGluLys 9

Db 26748 CAGATAATAACCAAAATAGTGAGAAG 26774

Search completed: August 6, 2003, 21:01:18  
Job time : 25470 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 6, 2003, 10:46:55 ; Search time 2381 Seconds  
(without alignments)  
91.869 Million cell updates/sec

Title: US-09-994-617-4  
Perfect score: 41  
Sequence: 1 QIEKRLIEK 9

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-O=/cpn2.1/uspto.spool.us09994617/runat\_29072003\_160423\_7651/app\_query.fasta\_1.199:  
-DB=EST\_QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09994617 @CGN.1.1.4237 -runat\_29072003\_160423\_7651 -NCPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_fod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	92.7	485	12	BG893927 kt21e06.y
2	38	92.7	793	29	BX207792 Danio rer
3	36	87.8	105	29	BZ769319 SALK_1419
4	36	87.8	300	9	AV177478 AV177478
5	36	87.8	360	14	C12277 C12277 Yuji
6	36	87.8	392	9	AU230851 AU230851
7	36	87.8	430	14	CB761156 AMGNNUC:S
8	36	87.8	441	13	BY286462 BY286462
9	36	87.8	443	9	AA047746 zf14f08.r
10	36	87.8	449	13	BX118247 BX118247
11	36	87.8	470	12	BM900614 IC39e12.y
12	36	87.8	480	12	BI324505 IK80H07.y
13	36	87.8	526	9	AU205517 AU205517
14	36	87.8	531	14	CB719039 AMGNNUC:M
15	36	87.8	551	12	BM902043 IC31e09.y
16	36	87.8	551	13	BM276124 BM276124
17	36	87.8	568	13	BM199076 BM199076
18	36	87.8	571	14	CB386125 OSTF039B6
19	36	87.8	573	12	BI863097 IM1e02.y
20	36	87.8	578	13	BM202105 BM202105
21	36	87.8	624	13	BM229293 BM229293
22	36	87.8	663	13	BM257525 BM257525
23	36	87.8	679	28	BH996315 Oef12b09
24	36	87.8	691	13	BM056791 BM056791
25	36	87.8	720	13	BM193980 BM193980
26	36	87.8	722	13	BM266869 BM266869
27	36	87.8	728	13	BM264007 BM264007
28	36	87.8	790	29	CC391111 PUH8X62TD
29	36	87.8	799	29	CC391106 PUH8X62TB
30	36	87.8	831	13	BH933817 AGENCOURT
31	36	87.8	834	28	AZ531187 ENTCC03TF
32	35	85.4	106	12	BM750362 K-EST0025
33	35	85.4	116	10	BF754856 MRO-CT045
34	35	85.4	262	9	AV088746 AV088746
35	35	85.4	292	10	BA448220 BA448220
36	35	85.4	322	28	AQ472163 CITBI-El-
37	35	85.4	373	13	BH789656 lo44c07.x
38	35	85.4	401	12	BM840791 K-EST0117
39	35	85.4	413	9	AI904742 CM-BT066-
40	35	85.4	417	9	AI474034 ti68e04.x
41	35	85.4	443	9	AI191919 GQ63a06.x
42	35	85.4	449	10	BG194864 RST14167
43	35	85.4	459	12	BI387942 BFL26_002
44	35	85.4	463	14	CD350047 UI-M-FY0-
45	35	85.4	470	10	BG185622 RST4573 A

ALIGNMENTS

RESULT 1  
BG893927  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BG893927 485 bp mRNA linear EST 04-JUN-2001  
kt21e06.y1 strongyloides ratti L1 PAMP1 v3 Chiapelli1 McCarter  
Strongyloides ratti cDNA 5' similar to TR:O16785 O16785 T21D12.4  
PROTEIN. [1] ; mRNA sequence.

BG893927  
BG893927.1 GI:14288537  
EST  
Strongyloides ratti  
Strongyloides ratti  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Panagrolaimidae; Strongyloidea; Strongyloides.

REFERENCE 1 (bases 1 to 485)  
 AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagaris,vili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE The Washington Univ. Nematode EST Project, 1999  
 JOURNAL Unpublished  
 COMMENT Contact: McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.  
 High quality sequence stop: 359.

# FEATURES source

1..485  
 /organism="Strongyloides ratti"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:34506"  
 /dev\_stage="L1"  
 /lab\_host="DH10B"  
 /clone\_lib="Strongyloides ratti L1 pMPL v3 Chiapelli McCarter"  
 /note="vector: pMPL (Gibco): The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dynal), PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pMPL1. Nematodes were provided by Dr. Mark Viney of Bristol, UK."  
 BASE COUNT 194 a 64 c 87 g 140 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 318 Length: 485  
 Score: 38.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 92.68% Indels: 0  
 DB: 12 Gaps: 0

US-09-994-617-4 (1-9) x BG993927 (1-485)

QY 1 GlnIleIleGluLysLeuIleGluLys 9  
 |||||.....:|||||  
 Db 249 CAATATTTCAGAAATTAATGGAAG 275

RESULT 2  
 Bx207792/c  
 LOCUS Bx207792 793 bp DNA linear GSS 13-MAR-2003  
 DEFINITION Danio rerio genomic clone DKEY-247K7, genomic survey sequence.  
 ACCESSION Bx207792  
 VERSION Bx207792.1 GI:28039678  
 KEYWORDS GSS.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 793)  
 Humphray,S.J., Huckle,E. and Durham,J.L.  
 Direct Submission  
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

COMMENT humquery@sanger.ac.uk Unpublished  
 This sequence was generated from the T7 end of BAC 247K7. 247K7 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:  
 http://www.sanger.ac.uk/Projects/D\_rerio/.

# FEATURES source

Location/Qualifiers  
 1..793  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-247K7"  
 /tissue\_type="Testis"  
 /note="vector pIndigoBAC-536"  
 BASE COUNT 272 a 111 c 129 g 281 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 597 Length: 793  
 Score: 38.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 92.68% Indels: 0  
 DB: 29 Gaps: 0

US-09-994-617-4 (1-9) x BX207792 (1-793)

QY 1 GlnIleIleGluLysLeuIleGluLys 9  
 |||||.....:|||||  
 Db 262 CAATATTTCAGAAATTAATACAAA 236

RESULT 3  
 Bz769319  
 LOCUS Bz769319 105 bp DNA linear GSS 13-MAR-2003  
 DEFINITION SALK\_141963.45.80.n Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK\_141963.45.80.n, genomic survey sequence.

ACCESSION Bz769319  
 VERSION Bz769319.1 GI:28943003  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 105)  
 AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Cadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.  
 TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

Unpublished  
 CONTACT: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu

COMMENT This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 3' end of At4g29010.

Class: TDNA tagged.  
 Location/Qualifiers  
 1..105  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_141963.45.80.n"  
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)

```

BASE COUNT      42 a   19 c   14 g   30 t
ORIGIN
Alignment Scores:
Pred. No.:      118      Length:      105
Score:          36.00      Matches:      7
Percent Similarity: 100.00%      Conservative: 2
Best Local Similarity: 77.78%      Mismatches: 0
Query Match:      87.80%      Indels:      0
DB:                29      Gaps:      0

US-09-994-617-4 (1-9) x BZ769319 (1-105)

QY      1 GlnllellleGluLysLeuileGluLys 9
Db      35 CAATCTCTACAGAGCTGATTGAGAAA 61

RESULT 4
AV177478/C
LOCUS   AV177478 300 bp mRNA linear EST 21-JUL-1999
DEFINITION AV177478 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo Caenorhabditis elegans cDNA clone yk529c5 3', mRNA sequence.
ACCESSION AV177478
VERSION   AV177478.1 GI:5557379
KEYWORDS EST.
SOURCE    Caenorhabditis elegans
ORGANISM  Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
AUTHORS   Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H., Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A., Mitani,Y., Iida,K., Desugi,H., Sugiyama,Y. and Nomoto,H.
TITLE     Expressed genes in C.elegans
JOURNAL   Unpublished
COMMENT   Contact: Yuji Kohara
          Genome Biology Lab.
          National Institute of Genetics
          Yata 1111, Mishima, Shizuoka 411, Japan
          Tel: 81-559-81-6854
          Fax: 81-559-81-6855
          Email: ykohara@lab.nig.ac.jp.
          Location/Qualifiers
            1..300
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk529c5"
            /sex="hermaphrodite"
            /dev_stage="embryo"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo"

BASE COUNT      71 a   72 c   53 g   103 t   1 others
ORIGIN
Alignment Scores:
Pred. No.:      453      Length:      300
Score:          36.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      87.80%      Indels:      0
DB:                9      Gaps:      0

US-09-994-617-4 (1-9) x AV177478 (1-300)

QY      2 llelleGluLysLeuileGluLys 9
Db      32 ATCATCGAAAGCTGATCGAAAAA 9

```

```

RESULT 5
C12277/c
LOCUS   C12277 360 bp mRNA linear EST 28-DEC-1998
DEFINITION C12277 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk148h11 5', mRNA sequence.
ACCESSION C12277
VERSION   C12277.1 GI:1559830
KEYWORDS EST.
SOURCE    Caenorhabditis elegans
ORGANISM  Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE     Expression map of the C.elegans genome
JOURNAL   Unpublished
COMMENT   Contact: Yuji Kohara
          Genome Biology Lab.
          National Institute of Genetics
          Yata 1111, Mishima, Shizuoka 411, Japan
          Tel: 81-559-81-6854
          Fax: 81-559-81-6855
          Email: ykohara@lab.nig.ac.jp.
          Location/Qualifiers
            1..360
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="CBI489 him-8(e1489)"
            /db_xref="taxon:6239"
            /clone="yk148h11"
            /sex="hermaphrodite, male"
            /tissue_type="whole animal"
            /dev_stage="varied"
            /clone_lib="Yuji Kohara unpublished cDNA"

BASE COUNT      95 a   69 c   48 g   138 t   10 others
ORIGIN
Alignment Scores:
Pred. No.:      572      Length:      360
Score:          36.00      Matches:      8
Percent Similarity: 88.89%      Conservative: 0
Best Local Similarity: 88.89%      Mismatches: 1
Query Match:      87.80%      Indels:      0
DB:                14      Gaps:      0

US-09-994-617-4 (1-9) x C12277 (1-360)

QY      1 GlnllellleGluLysLeuileGluLys 9
Db      153 CAATATAGAGAAATCTCATCGAAAAA 127

RESULT 6
AU230851
LOCUS   AU230851 392 bp mRNA linear EST 25-APR-2002
DEFINITION AU230851 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-93-A20 3', mRNA sequence.
ACCESSION AU230851
VERSION   AU230851.1 GI:19799561
KEYWORDS EST.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
AUTHORS   ; eurosoids II; Brassicales; Brassicaceae; Arabidopsis.
          Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
          Large scale analysis of Arabidopsis full-length cDNA

TITLE     Unpublished
JOURNAL

```

## COMMENT

Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4339  
Fax: 81-298-36-9060  
Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified paluscript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

## FEATURES

source

Location/Qualifiers  
1..392  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAFL19-93-A20"  
/tissue\_type="mixture of silique and flower"  
/lab\_host="DH10B"  
/clone\_lib="RAFL19"  
/note="Site\_1: BamHI; Site\_2: SalI; Subtraction Library"

## BASE COUNT

ORIGIN

166 a 65 c 77 g 84 t

## Alignment Scores:

Pred. No.: 638 Length: 392  
Score: 36.00 Matches: 7  
Percent Similarity: 100.00% Conservativeness: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 87.80% Indels: 0  
DB: 9 Gaps: 0

US-09-994-617-4 (1-9) x AU230851 (1-392)

QY 1 GlnlleleGlulysLeulleGluLys 9

DB 33 CAAATCTACAGAAGCTGATTGAGAAA 59

## RESULT 7

CB761156

LOCUS

CB761156 AMGNNUC:SRPB2-00116-H1-A srpb2 (10220) Rattus norvegicus cDNA clone  
srpb2-00116-h1 5', mRNA sequence.

ACCESSION

CB761156

VERSION

CB761156.1

KEYWORDS

EST.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 430)

Angen EST Program.

Angen Rat EST Program

Unpublished

Contact: Dan Fitzpatrick

Angen, Inc

One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881

Plate: 00116 row: h column: 1.

Location/Qualifiers

1..430

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/db\_xref="taxon:10116"

/clone="srpb2-00116-h1"

/tissue\_type="prostate tissue"

/clone\_lib="srpb2 (10220)"

/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; rat

prostate normalized double selected poly(A+) mRNA size

## COMMENT

fraction > 1 kb"  
BASE COUNT 111 a 56 c 101 g 51 t 111 others  
ORIGIN

## Alignment Scores:

Pred. No.: 718 Length: 430  
Score: 36.00 Matches: 7  
Percent Similarity: 100.00% Conservativeness: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 87.80% Indels: 0  
DB: 14 Gaps: 0

US-09-994-617-4 (1-9) x CB761156 (1-430)

## QY

1 GlnlleleGlulysLeulleGluLys 9

DB 249 CAGATCTTGAGAGACTGATAGAAAA 275

## RESULT 8

BY286462

LOCUS

DEFINITION

BY286462 RIKEN full-length enriched, visual cortex Mus musculus

ACCESSION

BY286462

VERSION

BY286462.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Okazaki.Y., Furuno.M., Kasukawa.T., Adachi.J., Bono.H., Kondo.S.,

Nikaido.I., Osato.N., Saito.R., Suzuki.H., Yamanaka.I., Kiyosawa.H.,

Yagi.K., Tomaru.Y., Hasegawa.Y., Nogami.A., Schonbach.C.,

Gojibori.T., Baldarelli.R., Hill.D.P., Bult.C., Hume.D.A.,

Quackenbush.J., Schriml.L.M., Kanapin.A., Matsuda.H., Batalov.S.,

Beisel.K.W., Blake.J.A., Bratt.D., Brusic.V., Chothia.C., Corbani

L.E., Cousins.S., Dalla.E., Dragani.T.A., Fletcher.C.F., Forrest

A., Frazer.K.S., Gaasterland.T., Gariboldi.M., Gissi.C., Godzik.A.,

Gough.J., Grimmond.S., Gustincich.S., Hirokawa.N., Jackson.I.J.,

Jarvis.E.D., Kanai.A., Kawaji.H., Kawasawa.Y., Kedzierski.R.M.,

King.B.L., Konagaya.A., Kurochkin.I.V., Lee.Y., Lenhard.B., Lyons

P.A., Maglott.D.R., Maltais.L., Marchionni.L., Mckenzie.L., Miki

H., Nagashima.T., Numata.K., Okido.T., Pavan.W.J., Pertea.G.,

Pesole.G., Petrovsky.N., Pillai.R., Pontius.J.U., Qi.D.,

Ramachandran.S., Ravasi.T., Reed.J.C., Reed.D.J., Reid.J., Ring

B.Z., Ringwald.M., Sandelin.A., Schneider.C., Semple.C.A., Setou

M., Shimada.K., Sultana.R., Takenaka.Y., Taylor.M.S., Teasdale

R.D., Tomita.M., Verardo.R., Wagner.L., Wahlestedt.C., Wang.Y.,

Watanabe.Y., Wells.C., Wilming.L.G., Wynshaw-Boris.A., Yanagisawa

M., Yang.I., Yang.L., Yuan.Z., Zavolan.M., Zhu.Y., Zimmer.A.,

Carninci.P., Hayatsu.N., Hirozane-Kishikawa.T., Konno.H., Nakamura

M., Sakazume.N., Sato.K., Shiraki.T., Waki.K., Kawai.J., Aizawa.K.,

Arakawa.T., Fukuda.S., Hara.A., Hashizume.W., Imotani.K., Ishii

Y., Itoh.M., Kagawa.I., Miyazaki.A., Sakai.K., Sasaki.D., Shibata

K., Shinagawa.A., Yasunishi.A., Yoshino.M., Waterston.R., Lander

E.S., Rogers.J., Birney.E. and Hayashizaki.Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-575 (2002)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Aizawa.K., Akimura.T., Arakawa.T., Carninci.P., Fukuda.S., Hirozane

T., Imotani.K., Ishii.Y., Itoh.M., Kawai.J., Konno.H., Miyazaki.A.

Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Michela Fagioli and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirotsawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

```
1. 441
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K53003C12"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
```

BASE COUNT 161 a 79 c 136 g 65 t  
ORIGIN

Alignment Scores:  
Pred. No.: 742 Length: 441  
Score: 36.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 66.67% Mismatches: 0  
Query Match: 87.80% Indels: 0  
DB: 13 Gaps: 0

US-09-994-617-4 (1-9) x BY286462 (1-441)

QY 1 GlnlIleGluLysLeuIleGluLys 9  
Db 309 CAGGTGGTGGAGAGACTATAGAAAAG 335

RESULT 9  
AA047746  
LOCUS  
DEFINITION  
IMAGE:zf14f08.rl Soares\_fetal\_heart\_NbHH19W Homo sapiens cDNA clone  
IMAGE:376935 5', mRNA sequence.

ACCESSION  
AA047746  
VERSION  
AA047746.1 GI:1527416  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 443)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

8889549

## PUBMED

## COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 502 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 277.

## FEATURES

source

```
1. 443
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1285191"
/db_xref="taxon:9606"
/clone="IMAGE:376935"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_fetal_heart_NbHH19W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(gt) primer [5',  
TGTTCACCACTCTGAAGTCGAGCGCGCATCTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBHH19W."
```

BASE COUNT 132 a 69 c 80 g 157 t  
ORIGIN

Alignment Scores:  
Pred. No.: 746 Length: 443  
Score: 36.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 87.80% Indels: 0  
DB: 9 Gaps: 0

US-09-994-617-4 (1-9) x AA047746 (1-443)

QY 1 GlnlIleGluLysLeuIleGluLys 9  
Db 10 AAAAGTATAGAAAAAATTATAGAAAAA 36

RESULT 10  
BX118247  
LOCUS  
DEFINITION  
IMAGE:BX118247 Soares\_fetal\_heart\_NbHH19W Homo sapiens cDNA clone  
IMAGE:998016871 ; IMAGE:376935, mRNA sequence.

ACCESSION  
BX118247  
VERSION  
BX118247.1 GI:27881271  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 449)  
Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M., Radelof, U., Schneider, D. and Korn, B.  
Human Unigeneset - RZPD3  
Unpublished  
Contact: Ina Rofes

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
RZPD; IMAGp998016871.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
Human Unigeneset - RZPD3 (RZPDLIB No.972)  
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/responderlibNo=972 Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
Heubnerweg 6, D-14059 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available royalty-free from RZPD;  
contact RZPD (clone@rzpd.de) for further information. Seq primer:  
M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

## FEATURES

source

1. .449  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE998016871 ; IMAGE:376935"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares.fetal.heart\_NbHL19W"  
/notes="Organ: heart; Vector: p7T3D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCAATCTGAAGTGGAGCGCCGATCTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Patima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NbHL19W."

BASE COUNT 139 a 69 c 79 g 160 t 2 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 759 Length: 449  
Score: 36.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 87.80% Indels: 0  
DB: 13 Gaps: 0

US-09-994-617-4 (1-9) x BX118247 (1-449)  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne hapla egg pAMP1 v1 Meloidogyne hapla cDNA 5'  
similar to TR:O16785 O16785 T21D12.4 PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:19383402  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne hapla  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

RESULT 11  
BM900614  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne hapla egg pAMP1 v1 Meloidogyne hapla cDNA 5'  
similar to TR:O16785 O16785 T21D12.4 PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:19383402  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne hapla  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

RESULT 12  
BM900614  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne hapla egg pAMP1 v1 Meloidogyne hapla cDNA 5'  
similar to TR:O16785 O16785 T21D12.4 PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:15003691  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne javanica (root-knot nematode)  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

RESULT 13  
BM900614  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne hapla egg pAMP1 v1 Meloidogyne hapla cDNA 5'  
similar to TR:O16785 O16785 T21D12.4 PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:15003691  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne javanica (root-knot nematode)  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

## COMMENT

Contact: McCartner JP  
The Washington Univ. Nematode EST Project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

The library was constructed by Claire Murphy and Dr. James McCartner  
at Washington University, St. Louis. DNA Sequencing by: Washington  
University Genome Sequencing Center  
Seq primer: -40RP from Gibco  
High quality sequence stop: 334.  
Location/Qualifiers  
1. .470

## FEATURES

source

/organism="Meloidogyne hapla"  
/mol\_type="mRNA"  
/db\_xref="taxon:6305"  
/dev\_stage="parasitic adult females"  
/lab\_host="DH10B"  
/clone\_lib="Meloidogyne hapla egg pAMP1 v1"  
/notes="Vector: pAMP1 (Gibco); Site\_1: NotI; Site\_2: SalI;  
The library was constructed by Claire Murphy and Dr. James  
McCartner at Washington University, St. Louis. The cDNA was  
made by using Dynabead oligo-dT priming (Dyna). PCR based  
library using a modified protocol from the SMART PCR cDNA  
Synthesis Kit from Clontech. Directionally cloned into the  
UDG sites of pAMP1."

BASE COUNT 204 a 53 c 76 g 137 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 805 Length: 470  
Score: 36.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 66.67% Mismatches: 0  
Query Match: 87.80% Indels: 0  
DB: 12 Gaps: 0

US-09-994-617-4 (1-9) x BM900614 (1-470)  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne javanica egg pAMP1 v6 Chiapelli McCartner  
Meloidogyne javanica cDNA 5' similar to TR:O16785 O16785 T21D12.4  
PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:15003691  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne javanica (root-knot nematode)  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

RESULT 12  
BM900614  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne javanica egg pAMP1 v6 Chiapelli McCartner  
Meloidogyne javanica cDNA 5' similar to TR:O16785 O16785 T21D12.4  
PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:15003691  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne javanica (root-knot nematode)  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

RESULT 13  
BM900614  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne javanica egg pAMP1 v6 Chiapelli McCartner  
Meloidogyne javanica cDNA 5' similar to TR:O16785 O16785 T21D12.4  
PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:15003691  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne javanica (root-knot nematode)  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

RESULT 14  
BM900614  
LOCUS  
DEFINITION  
rc39a12.y1 Meloidogyne javanica egg pAMP1 v6 Chiapelli McCartner  
Meloidogyne javanica cDNA 5' similar to TR:O16785 O16785 T21D12.4  
PROTEIN. [1] ; mRNA sequence.  
BM900614  
VERSION  
BM900614.1 GI:15003691  
KEYWORDS  
EST.  
SOURCE  
Meloidogyne javanica (root-knot nematode)  
ORGANISM  
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

McCartner, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

Email: est@watson.wustl.edu  
 The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.  
 High quality sequence stop: 383.

#### FEATURES source

```

1..480
  Location/Qualifiers
    /organism="Meloidogyne javanica"
    /mol_type="mRNA"
    /db_xref="taxon:6303"
    /dev_stage="enriched for eggs"
    /lab_host="DH10B"
    /clone_lib="Meloidogyne javanica egg pAMPl v6 Chiapelli McCarter"
    /note="Vector: pAMPl (Gibco); The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (dynamal). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMPl. Nematodes were provided by Dr. David Bird of North Carolina State University."
  BASE COUNT      210 a      55 c      82 g      133 t
  ORIGIN

```

```

Alignment Scores:
Pred. No.:      827      Length:      480
Score:          36.00     Matches:      6
Percent Similarity: 100.00%   Conservative: 3
Best Local Similarity: 66.67%   Mismatches:  0
Query Match:     87.80%     Indels:      0
DB:              12        Gaps:         0

```

US-09-994-617-4 (1-9) x B1324505 (1-480)

```

QY 1 GlnlleleGlulysLeulleGlulys 9
|||||:|||||:|||||:|||||
Db 433 CAAGTTATCCAAAATTAGTTGAGAAA 459

```

#### RESULT 13

```

AU205517
LOCUS AU205517 526 bp mRNA linear EST 17-JUL-2001
DEFINITION AU205517 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk847f05 5', mRNA sequence.

```

```

ACCESSION AU205517
VERSION AU205517.1 GI:14837888
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans

```

```

REFERENCE 1 (bases 1 to 526)
AUTHORS Kohara,Y., Shin-I,T., Thierri-Mieg,J., Thierri-Mieg,D., Suzuki,Y.
and Sugano,S.

```

```

TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

```

#### FEATURES source

```

1..526
  Location/Qualifiers
    /organism="Caenorhabditis elegans"
    /mol_type="mRNA"
    /strain="N2"
    /db_xref="taxon:6239"
    /clone="yk847f05"
    /sex="Hermaphrodite"
    /tissue_type="whole animal"

```

```

/dev_stage="L4"
/clone_lib="unpublished oligo-capped cDNA library, stage L4"
BASE COUNT      161 a      102 c      132 g      131 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      930      Length:      526
Score:          36.00     Matches:      7
Percent Similarity: 100.00%   Conservative: 2
Best Local Similarity: 77.78%   Mismatches:  0
Query Match:     87.80%     Indels:      0
DB:              9        Gaps:         0

```

US-09-994-617-4 (1-9) x AU205517 (1-526)

```

QY 1 GlnlleleGlulysLeulleGlulys 9
|||||:|||||:|||||:|||||
Db 421 CAAATCATCCAGAGCTGTTGAAAAA 447

```

#### RESULT 14

```

CB719039
LOCUS CB719039 531 bp mRNA linear EST 10-APR-2003
DEFINITION AMGNNUC:MRBEL-00001-f6-A RDS UNKNOWN (13985) Rattus norvegicus cDNA
clone mrbel-00001-f6 5', mRNA sequence.

```

```

ACCESSION CB719039
VERSION CB719039.1 GI:29776187
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

```

```

REFERENCE 1 (bases 1 to 531)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00001 row: f column: 6.

```

#### FEATURES Source

```

Location/Qualifiers
1..531
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="mrbel-00001-f6"
/cell_type="artificial construct"
/clone_lib="RDS UNKNOWN (13985)"
/note="artificial construct for unknown sequencing requests"
BASE COUNT      192 a      90 c      160 g      89 t
ORIGIN

```

#### Alignment Scores:

```

Pred. No.:      941      Length:      531
Score:          36.00     Matches:      7
Percent Similarity: 100.00%   Conservative: 2
Best Local Similarity: 77.78%   Mismatches:  0
Query Match:     87.80%     Indels:      0
DB:              14        Gaps:         0

```

US-09-994-617-4 (1-9) x CB719039 (1-531)

```

QY 1 GlnlleleGlulysLeulleGlulys 9
|||||:|||||:|||||:|||||
Db 107 CAGATCTTGAGAGACTGATAGAAAA 133

```

#### RESULT 15

```

BM902043
LOCUS BM902043 551 bp mRNA linear EST 12-MAR-2002
DEFINITION rc31e09.y1 Meloidogyne hapla egg pAMPl v1 Meloidogyne hapla cDNA 5'

```

similar to TR:O16785 O16785 T21D12.4 PROTEIN. [1] ;, mRNA sequence.  
BM902043  
VERSION BM902043.1 GI:19384831

EST

KEYWORDS  
SOURCE  
ORGANISM

Meloiodogyne hapla  
Meloiodogyne hapla

REFERENCE  
1 (bases 1 to 551)  
Tylenchoides; Heteroderidae; Meloiodogyninae; Meloiodogyne.

AUTHORS  
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, T.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

TITLE  
The Washington Univ. Nematode EST Project, 1999

JOURNAL

COMMENT

Unpublished

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Claire Murphy and Dr. James McCarter  
at Washington University, St. Louis. DNA Sequencing by: Washington  
University Genome Sequencing Center

Seq primer: -40RP from Gibco

High quality sequence stop: 405.

#### FEATURES

source

1..551

Location/Qualifiers

/organism="Meloiodogyne hapla"

/mol\_type="mRNA"

/db\_xref="taxon:6305"

/dev\_stage="parasitic adult females"

/lab\_host="DHI0B"

/clone\_lib="Meloiodogyne hapla egg pAMP1 v1"

/note="Vector: pAMP1 (Gibco); Site\_1: NotI; Site\_2: SalI;  
The library was constructed by Claire Murphy and Dr. James  
McCarter at Washington University, St. Louis. The cDNA was  
made by using Dynabead oligo-dT priming (Dynal). PCR based  
library using a modified protocol from the SMART PCR cDNA  
Synthesis Kit from Clontech. Directionally cloned into the  
UDG sites of pAMP1."

BASE COUNT 249 a 65 c 84 g 153 t

ORIGIN

#### Alignment Scores:

Pred. No.:	987	Length:	551
Score:	36.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	66.67%	Mismatches:	0
Query Match:	87.80%	Indels:	0
DB:	12	Gaps:	0

US-09-994-617-4 (1-9) x BM902043 (1-551)

QY 1 GlnIleIleGluLysLeuIleGluLys 9

||||:||||:||||:||||:||||

Db 321 CAAGTTATTCAAAAATAGTGA AAAA 347

Search completed: August 6, 2003, 12:55:58  
Job time : 2400 secs